Planning Report 03-1

Methods for Assessing the Economic Impacts of Government R&D

Gregory Tassey
Senior Economist

National Institute of
Standards & Technology

Program Office
Strategic Planning and
Economic Analysis Group

September 2003

NIST
U.S Department of Commerce
Technology Administration

Methods for Assessing the Economic Impacts of Government R&D

Gregory Tassey

Senior Economist
National Institute of Standards and Technology

tassey@nist.gov

http://www.nist.gov/public_affairs/budget.htm

September 2003

Abstract

Analyses of the actual or potential economic impacts of government R&D programs have used a number of distinctly different methodologies, which has led to considerable confusion and controversy. In addition, particular methodologies have been applied with different levels of expertise, resulting in widely divergent impact assessments for similar types of R&D projects. With increased emphasis on government efficiency, the current state of methodology for strategic planning and retrospective impact analyses is unacceptable.

NIST has over the past decade conducted 30 retrospective microeconomic impact studies of its infratechnology (laboratory) research programs. Additional microeconomic studies have been conducted of technology focus areas in its Advanced Technology Program (ATP) and of the aggregate impacts of its Manufacturing Extension Partnership (MEP) Program. In addition, NIST has undertaken prospective (strategic planning) economic studies of technology infrastructure needs in a number of divergent and important industries. From these studies have evolved methodologies for conducting microeconomic analyses of government technology research and transfer programs.

The major steps in conducting economic impact studies are identifying and qualifying topics for study, designing an analytical framework and data collection plan, conducting the empirical phase of the study, writing a final report and summaries of that report, and disseminating the results to government policy makers, industry stakeholders, and other interested parties.

Execution of these steps is not straightforward. No consensus exists with respect to scope and depth of industry coverage, development of an analytical framework (including choice of metrics and impact measures), and design of data collection strategies. Even when an acceptable methodology is chosen and effectively executed, the results are frequently not understood by policy makers. NIST has therefore developed a methodology over the past decade that addresses the technology-based economic activity being studied, is appropriate for the nature of the government program or project responding to an underinvestment phenomenon, and provides an analysis understandable by industry and government stakeholders.

Based on the NIST experience, this report describes methodologies appropriate for economic impact assessments of Government R&D programs and gives numerous examples of their application to specific studies. Guidelines for interpretation of both qualitative and quantitative results are provided.

Table of Contents

Tables

Figures

Methods for Assessing the Economic Impacts of Government R&D

Gregory Tassey

In spite of efforts in the United States over the past decade to implement standardized methodologies for assessing the economic impacts of government research programs, no generally accepted approach yet exists. Program managers therefore have to rely on either external consultants who tend to emphasize one particular technique over others or on experimentation by agency analysts with alternative assessment methods in the hope of evolving a suitable methodology over time that is also accepted by external audiences.

In fact, a single "manual" for impact assessment may never be achieved. The technology trajectories and economic outcomes that government programs or projects seek to leverage vary significantly, as do the complex economic structures that characterize a technology-based economy.[1] Thus, no single metric or measurement method can (1) address the diversity and complexity of an R&D agency's technological outputs, (2) describe the subsequent processes by which private sector impacts occur, and then finally (3) accurately capture the resulting economic outcomes.

---

[1] The terms "program" and "project" are used regularly throughout this report. A research *program* is defined as a major ongoing activity targeted at providing long-term support for a particular element of an industrial technology. Thus, it responds to a systematic generic market failure, which is expected to exist for some time and affect a range of related economic activity. Unfortunately, the term is used at two distinct levels. One is economy-wide (example: NIST's Advanced Technology Program (ATP), which targets the first (generic) phase of technology research). The other level is a specific technology, whose development requires long-term support (example: DoE's fuel cell research program). A research *project* is a specific implementation of a research program's mission, focusing on, say, a specific technology element within a specific technology life cycle. An example is a project to develop a measurement method for testing semiconductor chips within an ongoing program of providing measurement infratechnology support to the semiconductor industry.

However, certain methodological elements are common to all economic impact assessments. These include

(1) decision criteria for selecting programs/projects for study,

(2) development of impact scenarios and subsequent metric construction,

(3) development of impact hypotheses,

(4) identification of primary and secondary data sources,

(5) design and conduct of data collection,

(6) analysis of the data, and

(7) compilation of the analysis into a final report that effectively communicates essential impact information to the intended audiences.

This report reviews alternative analytical frameworks, metrics, data collection strategies, and impact measures that summarize the values obtained for the selected metrics in order to implement the above seven steps. The report then discusses interpretation options and suggests some overall guidelines for assessing the economic impacts of government technology research programs.[2]

1.0 Policy Motivations for Doing Economic Impact Assessments

For decades, R&D agencies allocated resources largely in an unstructured process. Retrospective impact assessments were infrequent and ad hoc and were usually motivated by an externally imposed directive. In the past decade, this situation has begun to change, as government efficiency concerns have increased.

1.1. The Imperative to Assess Government R&D Programs

The emerging demand for economic impact assessments of government research programs is the result of relentless growth in global competition. In the United States, gross domestic product (GDP) in real terms (adjusted for inflation) has grown 121 percent over the past 25 years (1977–2002). In the same period, industry-funded R&D has grown 159 percent in real terms, indicating the increased dependence by industry on technology for competing in global markets. Yet, although technology (and hence the R&D process that produces it) increasingly is recognized as the main driver of long-term productivity growth, total factor productivity (TFP) during this period (1976–2001) has grown at a slow average annual rate of 0.53 percent. In recent years, this contradiction has lead to scrutiny of both the amount and the composition of R&D investment. Government's funding and conduct of R&D have received even more attention, as its

[2] Much of the analysis presented in this report is based on the extensive experience of the National Institute of Standards and Technology (NIST) with studies of the economic impacts of technology research programs. During the past decade, NIST has conducted over 40 prospective and retrospective economic studies. NIST has a diverse set of technology support missions (programs) and the experience gained has resulted in a steady advance in understanding, selecting, and executing analytical methodologies appropriate for assessing the economic impacts of a range of government R&D programs.

share of national R&D funding has declined and the philosophical and economic rationales for its roles have been the subject of increasing debate.

Analysis of government R&D investment has been stimulated further by increasing efforts in the 1990s toward greater efficiency in government in general, highlighted by the passage in 1993 of the Government Performance and Results Act (GPRA). Over the ensuing decade, both the executive and legislative branches have steadily increased pressure on Federal agencies to undertake economic impact assessments. In GPRA terminology, this mandate has meant collecting data and estimating project "outputs" and "outcomes". The former is the direct technical output of a research program and the latter is the impact on the ultimate target segment of society. Similarly, new funding requests are required to be accompanied by economic impact projections.

GPRA is designed to promote regular (yearly) reporting against goals using a limited set of generic and highly measurable indicators. Unfortunately, this framework has restricted its utility as an effective impact assessment and planning tool. Typical research projects take many years to complete and therefore produce economic impacts ("outcomes") after considerable time has elapsed. Moreover, several different economic impacts are possible from the same project, which are typically spread out in time and generate different types of technical output. Such patterns confound the GPRA requirement for annual reporting against the same metric. Interim technical "output" measures can be compiled, but these metrics suffer from the intermittent occurrence problem and cannot substitute for the outcome measures upon which the project's ultimate success or failure depends.

In response, the Federal Government has begun to implement systematic planning and impact assessment policies that require more detailed reporting and targeted metrics. White House officials in the Bush Administration initiated a broad management improvement agenda, which included initial development in 2001-2002 of guidelines for government investment in both basic scientific research and applied (technology) research.[3] The President's science advisor stated the overall issue as

> *"Scientists do, of course, make judgments all the time about promising lines of research.... It makes sense for the world's largest sponsor of research, the U.S. Government, to want to make such choices as wisely as the most productive scientists do... But is it possible to decide rationally when to enhance or to terminate a project if we do not possess a way of measuring its success?"*
>
> John Marburger (keynote speaker at the American Association for the Advancement of Science's 27th annual Colloquium on Science and Technology Policy, 2002)

Central to this strategy, the Office of Management and Budget (OMB) developed a program review algorithm called Program Assessment Review Tool (PART) to

[3] White House memorandum on "FY 2004 Interagency Research and Development Priorities", May 30, 2002.

implement the established criteria for government programs. PART has been applied to specific R&D programs beginning in 2002. However, the required tools for complete and effective management of government R&D programs are not yet available. Many factors explain this inadequacy, not the least of which is the general lack of analytical capability in government R&D agencies and the failure to fund methodology development research in universities. Thus, the evolution of economic impact assessment has been slow and uneven. PART, by allowing more varied and program-specific impact indicators has the potential for a more in depth and focused assessment, but it still relies on the R&D agency's capability to supply the range of impact metrics and data that allow accurate and convincing impact assessments to be made.

At one extreme, economic impact assessment may be undertaken only occasionally in response to external pressure. At the other extreme, assessments may be an institutionalized process with a number of both retrospective and prospective studies undertaken each year, the results of which are then used systematically in program evaluation and resources allocation.

Currently, pressures to conduct systematic strategic planning and retrospective impact assessments of research programs/projects are of relatively recent vintage, so most agencies have not acquired the internal capability to select appropriate impact metrics, data sources, and analytical methods or to select external contractors with the appropriate impact assessment skills. Moreover, R&D agencies are for the most part managed by technically trained people who are unfamiliar with economic assessment tools and either have difficulty understanding the imperative for such analysis or are uncomfortable with the use and interpretation of information produced by a distinctly different discipline. Finally, while some universities have curricula that include impact assessment techniques, little of it is designed for government research program evaluation.[4]

1.2. Need for Economic Analysis

Most government research programs either target economic impact as the final outcome or as the means to achieve a social objective. Thus, implementing guidelines for the management of resource allocation in government research programs requires a set of analytical and empirical tools to

(1) Identify the elements of technologies that warrant government support (i.e., suffer from systematic underinvestment by the private sector);

(2) Enable R&D agencies to construct metrics that reflect the technical outputs and economic outcomes of their respective missions;

[4] Public policy programs with a few exceptions develop and apply impact assessment methods to a range of social programs and largely ignore technology. Business schools have well-developed metrics and measures of R&D project impact, but the focus is on the corporate R&D program or project. Both areas of expertise have only partial relevance for government R&D programs. Some of the few efforts directed at government R&D impact assessment include Kostoff's [2001] review of the literature for assessing the impact of scientific research and reviews by Tassey [1996, 1999] and Link and Scott [1998] of approaches for analyzing the economic impacts of government-funded technology research.

(3) Gather comprehensive and accurate data from industry on private investment trends, and construct policy-relevant analyses from the metrics and data to guide resource allocation[5]; and

(4) Gather comprehensive and accurate data from industry on the impacts of ongoing or completed R&D programs.

1.3. Managing the Technology-Economic Impact Interface

In the past, the budgets of federal R&D agencies have been determined by general and largely subjective information. Neither the size nor especially the content of an agency's budget has reflected systematic analysis of expected payoffs among programmatic alternatives. Missions that are focused on visible and bounded goals tend to do better in the budget process than those that are more diffuse and hence less visible and more difficult to understand. As a result, large and expensive technology programs often have been established within focused mission agencies without a substantive strategic planning effort and have remained in place for long periods of time without periodic impact assessments.

In the future, the trends towards more analysis to support decision making mean that all agencies will face the considerable challenge of conducting planning and evaluation at several levels. These levels range from the overall sector or mission rationale to the allocation of resources across specific programmatic areas and, finally, projects within those areas.

R&D agencies with diffuse missions will have more difficult planning and impact assessment challenges because of the broader scope of technologies and associated industry structures that must be assessed. Several factors confound effective resource allocation for these R&D support missions that transcend technologies and industrial sectors. First, agencies with missions that require funding a range of technologies that support a social objective (energy independence or health care) or with missions that support technical infrastructure that cuts across industry and technological boundaries (measurement infratechnologies) require considerably more resources for effective strategic planning. The breadth of an R&D agency's mission creates a severe portfolio management problem. Even large companies with multi-billion dollar R&D budgets have a distinct technological focus to their R&D portfolios. These large companies have R&D foci that only target market applications derived from segments of generic technologies such as computer hardware, computer software, pharmaceuticals, etc.[6]

Second, programs such as those supplying technical infrastructure are particularly difficult to manage because they lack visibility to stakeholders and the research outputs are neither an explicit input nor an output in economic activity. However, the diffuse

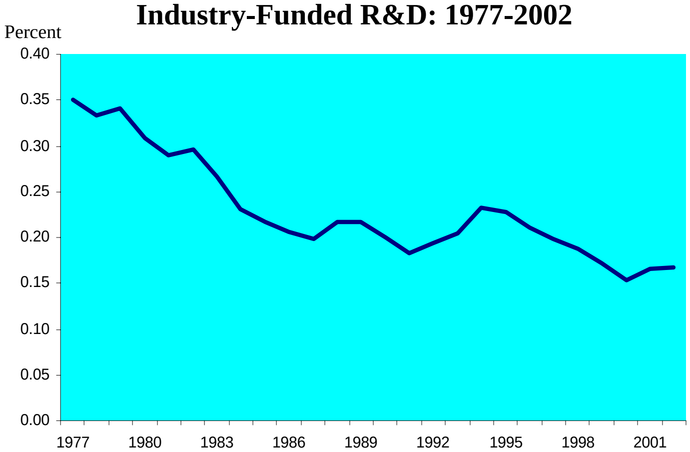
---

[5] The OMB criteria can be found at http://www7.nationalacademies.org/gpra/. For a discussion of the criteria see http://www7.nationalacademies.org/gpra/Basic%20Research.html.

[6] The National Institute of Standards and Technology (NIST) is an extreme example of an R&D agency with a diffuse mission. NIST's mission is to provide several types of technical infrastructure support to U.S. industry broadly, which means managing a portfolio of research projects that potentially address any technology.

Figure 1

Ratio of NIST Laboratory Funding to Industry-Funded R&D: 1977-2002



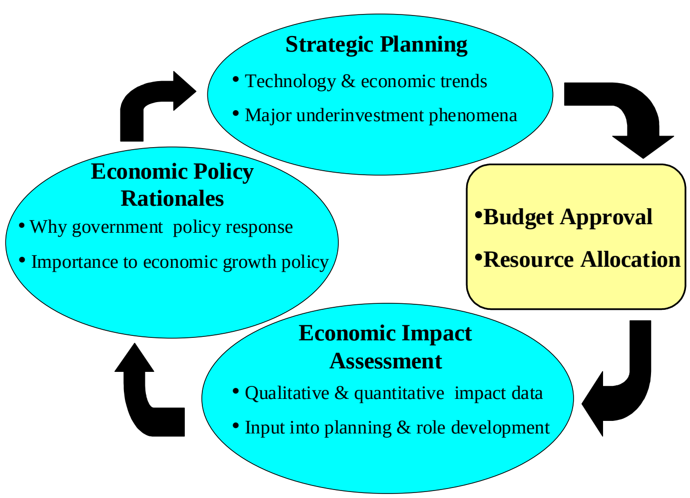
Source: National Science Foundation; NIST Budget Office

character of some categories of technical infrastructure does not imply less economic impact. In fact, the reverse can be true. Over 30 retrospective economic impact studies of NIST laboratory research have been conducted.[7] The results indicate large payoffs from a wide range of outputs in this category of technology infrastructure and associated standards. Yet, the agency's laboratory research budget has grown 92 percent in real terms over the 25-year period during which industry-funded R&D (the primary target of NIST support) grew 159 percent. The difference in growth rates between industry-funded R&D and funding for NIST laboratory research implies a significant *relative* decline in NIST's capabilities over this period, as indicated by the ratio of the two funding trends in Fig. 1.

This relative decline could be due to any of three reasons: (1) the policy process overestimated the NIST infrastructure support role compared to private sector investment in technology in the early portion of this time period and subsequently corrected the excess resource allocation; (2) a substantial reduction in the need by industry occurred over time for NIST's infratechnologies and related standards services, so the declining NIST role simply reflects effective policy adjustment; or, (3) the decline represents a failure by NIST to jointly identify and prioritize needs with industry and then demonstrate economic impact, both prospectively and retrospectively.

Figure 2

Economic Analysis for Government R&D Programs



With respect to the third possibility, NIST has devoted considerable resources to economic impact assessment over the past decade, particularly for retrospective impact studies. The results from these studies have been impressive and studied by other government R&D agencies in the United States and other countries.

[7] See the Appendix for a listing of these studies and summary information.

However, this agency's budget experience emphasizes the need to adopt an even more comprehensive approach. Such an approach includes more prospective or strategic planning studies and economic role assessment. The last category utilizes the retrospective and prospective study results to support analyses explicitly rationalizing government intervention in private markets.[8]

Fig.2 shows that economic analysis has three distinct uses in the management of government R&D projects. The initial and most challenging application is the development of economic rationales for a government role in supporting private sector investment in technology. Without a fully developed and articulated rationale, the program either will not be approved or will be under constant attack and will likely proceed with limited resources. If an economic rationale is accepted, implementation requires effective strategic planning.[9] If the economic analysis supporting strategic planning shows systematic and significant underinvestment, funds should be approved for a government research program targeting the underinvestment. Finally, increasing concerns over the past decade with respect to government efficiency have promoted more retrospective impact assessment to determine how ongoing projects an ultimately programs have performed. The results of these impact studies also feed back into the economic rationale activity, thereby completing the circular flow of economic assessments indicated in Fig. 2.

In summary, the motivations for conducting economic impact studies of government research programs are threefold:

(1) Economic studies collectively provide a database of the nature and magnitude of economic impacts (outcomes) from the research supported (outputs). Such a database is increasingly required by the policy and budget arenas and by an R&D agency's industry stakeholders and advisers. The policy arena has emphatically made the point that demonstration of economic impact is essential for funding budget initiatives and, in fact, for continuance of existing programs.

(2) If conducted correctly, these studies provide input into reassessments and better articulations of an R&D agency's roles. The policy process is constantly debating the scope and size of specific government technology support functions. Therefore, an ability to articulate economic roles in concrete terms—specific rationales, types of programmatic responses, mechanisms of delivery, and finally magnitudes of economic impacts—is essential for the continued existence of the basic mission.

(3) The increasing imperative to improve management of government research programs requires more and better data on the expected impacts of proposed or ongoing programs in order to assist the process of determining priorities and designing implementation, ongoing management, and technology transfer strategies.

[8] See Tassey [forthcoming] for an R&D underinvestment analysis framework.

[9] Section 5 provides a summary of prospective (strategic planning) economic study.

2.0 Uses of Economic Analysis

This section describes the elements of the basic framework for analyzing the economic impacts of completed, existing, and prospective government-funded projects whose objectives are to advance public good elements of industrial technologies. Based on this framework, subsequent sections will provide more specific discussions of metrics, data collection strategies, and interpretation techniques.

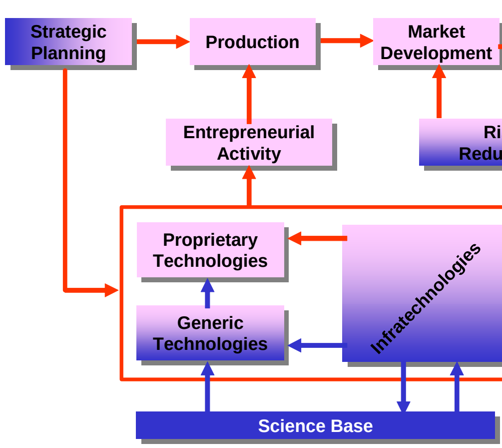
2.1. The Rationale for Government Intervention in Private Markets

Most government R&D programs are established without a systematic assessment of the scope and magnitude of the needed intervention. This complicates both planning and subsequent impact studies and, in fact, can lead to protracted policy debates and even embarrassing results in the form of either no impact or unintended negative impacts. Thus, the initial use of economic impact assessment should be to provide information on the appropriateness of a government role in supporting the evolution of an industrial technology. As a result, both the R&D agency and the analyst should view impact studies within the context of the economic rationales for the research program's existence. The rationales for government funding of research to support economic growth are derived from what economists call "market failure" arguments (Tassey [forthcoming]). These rationales are based on identification, characterization, and measurement of barriers to private-sector investment in R&D, where the term "barrier" implies underinvestment in either amount or composition of the R&D.

Tassey [1997, forthcoming] provides a framework for identifying and characterizing R&D underinvestment phenomena. The typical industrial technology is disaggregated into several major elements characterized by significantly different infrastructure (public good) content and hence distinctly different sets of investment barriers. The existence of different investment barriers is the key construct in determining government R&D support roles. This disaggregation is shown in Fig. 3, in which the shading indicates the degree of public good content in each of the major elements of the typical industrial technology. The technology box at the bottom is derived from an underlying science base. The existence of several distinct elements comprising industrial technologies defies the notion that technologies are "black boxes" that emerge from the science base as a homogeneous entity and directly enter a production process. Instead, the three technology elements shown arise from different sources in response to distinctly different investment incentives and research processes.

Figure 3

Economic Model of a Technology-Based Industry



Source: G. Tassey, *The Economics of R&D Policy*, Quorum Books, 1997, p. 70

Specifically, an industrial technology is based on a set of fundamental or generic concepts. Although examples can be found of technologies emerging before significant proof of concept, an industry's *generic technology* increasingly must evolve (basic concepts demonstrated, prototypes developed and tested) before industry is willing to commit significant funds to the more applied R&D required for market applications of the technology.

This evolutionary nature of technology development and commercialization is indicated in Fig. 3 by the arrows showing the direction of progressive knowledge application from basic science to generic technology development to proprietary products, processes, and services. Further, the diagram indicates that this evolutionary process (which is more complicated than shown because of feedback loops) is facilitated and in some cases made possible by a set of technical tools called *infratechnologies*. These tools (including measurement and test methods, technical support for interfaces between elements of systems technologies, scientific and engineering databases, and techniques such as quality assurance procedures) are ubiquitous in technology-based industries (often exerting their impacts the form of industry standards).

Which technology element is the target of the government research program/project determines the analytical and data collection approaches to strategic planning and retrospective impact assessment. Assuming the target has been determined by underinvestment analysis, the analyst will choose an analytical framework based on the type of technology infrastructure targeted. Doing so will allow accurate determination of the nature of the prospective/retrospective technical outputs from the research, the specific outcome (economic impact) metrics to be estimated/measured, the relevant types of qualitative analyses of the impact, and summary economic role assessments that will provide feedback/justification to government managers and other stakeholders (in particular, industry and the Congress).

For retrospective impact assessments, the primary focus is the project, with program level assessment issues discussed as appropriate. Prospective studies are usually broader in scope because their purposes are to first identify major areas of need for government R&D support and then to provide information that helps select projects to implement the program's objectives.

In general, for R&D agencies with missions that include the support of industrial growth, economic metrics should dominate planning and impact assessment. This imperative includes the majority of R&D agencies, even those with focused missions (health care, energy independence, environment protection). The reason is that executing these missions requires R&D agencies to be concerned with the evolution of industry structure and investment behavior that deliver the mission-critical technologies and related services.

2.2. The Analytical Framework for Economic Impact Assessments

Economic impact studies in R&D agencies require a multidisciplinary focus. Specifically, they should use the disciplines of technology assessment, microeconomic analysis, and corporate finance to provide an acceptable framework for the desired analysis.

2.2.1. Project Selection and Study Objective Statement. As a preliminary step in the application of economic analysis, selection criteria should be applied to determine what projects will be studied (retrospective) or what candidate technological areas will be assessed (prospective). The nature and use of study selection criteria will vary significantly depending on the level of commitment to planning and project impact assessment.

All retrospective impact assessment studies should begin with a careful analysis of objectives to avoid misinterpreting the project's economic role. Such a misinterpretation results in poor selection of impact measures. Specifically,

- In the case of retrospective impact assessments, provide a characterization and assessment of the research program's objectives and summarize the public policy (market failure) rationales for undertaking the specific project;
- In the case of prospective impact studies, identify and characterize the nature and magnitudes of the suspected market failure (private sector underinvestment) that would lead to the need for government-funded research.

For existing research projects, the objectives reflect the perceived market failure that rationalized their creation within the context of the broader rationale for the parent program. Ultimately, the retrospective analysis should provide decision makers with information that helps reassess the original market failure rationale for establishing the research program. Thus, retrospective analyses are not just a mechanism for generating performance scorecards. They are truly management tools that can help to significantly alter management of future research programs or, if undertaken in mid-stream, even adjust management of individual projects or portfolios of related projects.

For strategic planning (prospective) studies, the objective is to develop substantial quantitative and qualitative information that illuminates the scope and magnitude of a set of related market failures affecting an industry or supply chain and thereby facilitate resource allocation at the program level.[10] Hence, the focus is on characterizing technical barriers and estimating the cost to the economy of not removing them.

If economic impact assessment is part of a broader program evaluation effort, projects might be chosen at random or scheduled for periodic review. However, given the current lack of resources dedicated to impact assessment in most R&D agencies, taking a "target-of-opportunity" approach is more practical. In the latter case, the objective is to select projects for which preliminary screening indicates that substantial market impact has occurred or is likely to occur in the near future. The latter approach has several implications for the resulting database:

---

[10] The term "supply chain" refers to a set of industries that are virtually integrated vertically. Each level (industry) in a supply chain adds value until final demand (a product or service) is met. The sum of the value added by each level is the supply chain's contribution to GDP. An example of a first level in a supply chain would be silicon and other semiconductor materials. These materials are used to manufacture semiconductor devices, which are combined to form electronic components and equipment such as computers. The latter are further combined with other categories of equipment to form "systems," such as an automated factory that manufactures a product (computer) or a telecommunications network that provides a service.

- A cross-section of technologies and industries studied can be achieved over time with a modest expenditure of funds, but the results can nevertheless provide a balanced perspective of the nature of the R&D agency's economic impacts and the range of potential impact for successful projects.

- Less-effective projects in general will not be examined, but the lessons learned from a broad range of selected projects will improve technical managers' understanding of project selection and design without utilizing substantial resources on projects that during an initial screen are judged to have not attained significant payback;[11]

- The resulting database can provide the policy and budgets arenas, as well as industry constituents, with a perspective on the target range of qualitative and quantitative economic impacts of the R&D agency's mission and hence on the original rationale for the agency's role.

Effective screening of study candidates typically requires one or more meetings between the agency's economics and technical staffs to effectively apply the selection criteria. Specifically, candidate impact studies are screened to

(1) determine to the extent possible if the project has had substantive economic impact,

(2) assess the feasibility of successfully carrying out a study, and

(3) accurately estimate the resources required to complete the analysis.

2.2.2. Timing of the Study. The ideal time to conduct an economic impact study seems to be about three to ten years after significant marketplace impact has commenced. The reason for requiring at least a three-year waiting period is that some minimum time is needed for the generic technology or the technological infrastructure affected by the R&D agency to diffuse widely. Only then can significant economic impacts be realized within the benefiting industries and observed and estimated by industry respondents to a survey.

The ten-year maximum waiting period is determined by the fact that, as time elapses, the analyst typically begins to encounter difficulty in identifying and locating industry managers with substantive knowledge of the project, so that accurately assessing the project's full range and magnitude of impact is compromised. The latter problem can occur even if the actual impact period is longer than ten years. In latter case, economic benefits can be extrapolated forward in time from the higher quality impact data obtained in the optimal study period. The case studies undertaken by NIST beginning in the early 1990s show a wide variance in market diffusion patterns in terms of length of impact and timing of maximum impact, so the three-to-ten-year interval is an approximate guideline.

---

[11] However, the NIST experience indicates that technical managers do not always have an accurate perspective on the relative impacts of projects that make up a research program. As a result, some percentage of projects recommended for study will be "low-yield" ones and, if fully analyzed, will thereby provide direct information on failure modes. Even good screening techniques can fail to excise all such candidates. The opposite possibility—overlooking high-impact projects—is less likely, but not impossible.

A number of the NIST economic impact studies demonstrate the difficulties in examining programs that have existed for several decades. Studies of such diverse NIST research programs as antenna metrology, data encryption standard (DES) for electronic funds transfer, and cholesterol measurement indicate that the early projects within these programs had large economic impacts after initial transfer of infratechnology began (the 1980s for antenna metrology and the 1970s and 1980s for both DES and cholesterol measurement). All three programs have continued to provide infrastructure support to industry through subsequent complementary projects for several decades.

However, attempts to estimate the economic impacts of these three programs encountered considerable difficulty due to both conceptualization and data collection problems associated with the long time periods. In the antenna metrology study, for example, the analysts had considerable difficulty gathering both benefit and cost data, as well as defining at what point in time the study period should begin. As a result, they were able to construct only limited net benefit estimates. In the cholesterol study, analysis of the early years of the program (probably when maximum impact occurred) was not even attempted.[12]

The implication is that including the earliest years of a research program's life cycle is highly desirable because of the frequent large impact of the technology infrastructure that occurs early in this time frame relative to other contributions from follow-on projects (calibration services, standard reference materials, technical consulting and other mechanisms of technology transfer). For example, progress in improving cholesterol accuracy over the entire 30-year period of NIST involvement can arguably be traced to NIST's development of the definitive measurement method for cholesterol measurement, isotope dilution mass spectrometry (IDMS). In this impact study, the analyst concluded that going back in time to the first part of this period was not feasible, so the impact of the original method was not accounted for directly (in fact, it was considered a sunk cost). The analyst then limited the scope of analysis to the subsequent "measurement system" implementations of the basic method.

In contrast, a study of another program in the same NIST laboratory, standard reference materials (SRMs) for sulfur in fossil fuels, captured the impact of the development and use of the same generic method (IDMS). Estimates of this impact were included with follow-on impact estimates of IDMS' application to pollution measurement activity, extending the study period and aggregate economic impacts estimated. The latter study subsequently produced very large impact estimates. This difference in study periods and consequent ability to obtain more impact data is often a factor in explaining the substantial differences in aggregate economic impact estimates across studies.

Finally, this emphasis on the importance of projects that produce generic measurement infratechnologies early in the life cycles of research programs does not mean that the typical program's life should be shorter. The initial infrastructure development (usually a method) requires follow-on projects to achieve widespread and efficient transfer to industry.[13] Thus, without these latter "tech-transfer" projects, much of

[12] See Leech [2000] and Leech and Chinworth [2001].

[13] For example, in the case of infratechnologies, the development of a new test method or process model is typically followed by the development of calibration methods and standard reference materials along

the economic benefits from the initial infratechnology may not be realized. However, the NIST impact analyses imply that these follow-on projects might be either smaller in size or truncated in time, as discussed in following sections of this report.

2.2.3. Background Analysis and General Analytical Approach. Early in the impact assessment process, the nature, scope, and roles of the relevant technology and the supporting technological infrastructure must be mapped out and related to the relevant industries and the competitive dynamics of the associated markets. This step is necessary for effective subsequent development of the hypothesized economic impacts. Such background or economic context analysis requires an applied microeconomic and industrial organization approach, so that technology trends, corporate strategies, and any external influences such as regulations can be combined into a context for understanding the role of the technology support program or project being studied.

Identifying the "relevant" industries is particularly important because the selection will determine the population to be surveyed. Clearly, government research support will have direct and usually major impacts on the industries that develop the technology and on those industries that buy the technology. Analyzing this limited set of industries enables impact assessments that specifically relate the public and private R&D investments to significant market impact. Economic impacts clearly extend to other industries and eventually to the final users in the relevant supply chains. However, the relationships between the original R&D investment and its impacts in these downstream industries become increasingly blurred as other investments (including the R&D in those industries) must be added to the set of explanatory variables.

R&D agencies have often succumbed to the temptation to apply macroeconomic impact assessment techniques in order to include the impacts on these other industries. Such analytical approaches attempt to sum up a ripple (multiplier) effect across the economy from an intervention in a limited number of private markets. Input-output models of the economy are used to obtain the estimates. Each industry' interaction with other industries is represented by a series of coefficients, which represent that industry's dependence on or contribution to other industries. These relationships or "multipliers" produce large impact estimates when added up across an entire economy.

Unfortunately, several problems with this approach argue against its use. An early well-publicized case in point was NASA's attempt in the 1970s to estimate the national economic impact of its entire set of R&D programs. NASA hired an economic consulting firm, Chase Econometrics [1975], to estimate the impact of NASA's R&D on aggregate total factor productivity (TFP) growth. Chase conducted a time series analysis (14 to 18 years) of NASA's R&D and "other" R&D expenditures in the economy and related these expenditures to TFP. Very large impact estimates were produced. The study was criticized by economists for the obvious reason that relating relatively modest R&D expenditures that are focused on specific technological areas (and hence specific areas of

---

with critically evaluated technical data to be used by industry for implementing the new method. Improvements in the basic method also occur before the technology life cycle being supported comes to an end.

the economy) to the entire national economy was not justifiable, given the relatively crude disaggregation of national economic activity in input-output models.[14]

More recently, NIST's Advanced Technology Program (ATP) experimented with the use of input-output analysis to estimate the national economic impact of a joint research program with a consortium of companies in the automotive supply chain (CONSAD [1997]). A small research expenditure of approximately $30 million (in an economy at the time spending over $200 billion on R&D) was related to macroeconomic impact variables (GDP, employment). Again, very large impact estimates were obtained. However, the cells in an input-output matrix are large industry groups. Therefore, such a small interjection of funds is barely noticeable within the cell in the input-output matrix (transportation equipment) to which the R&D is directed.[15] Many factors simultaneously affect the outcome measures (value added, employment) for a large sector, which are not distinguishable in such a model; that is, attribution to such a small injection is beyond the sensitivity of the model.

National impact estimates can be estimated for particular technologies or technology infrastructure under certain conditions. To do so requires a microeconomic analysis that acquires comprehensive impact data on the investment or activity and subsequent impacts in a number of industries, and then provides reasonable assumptions about similar impact mechanisms in other economic sectors. This approach has the advantage of explicitly estimating the economic impacts of the intervention (or, in prospective studies, the underinvestment) and utilizing the elements of the relevant microeconomic activity for extrapolation. If the scope and depth of this analysis is sufficient to provide a reasonable level of confidence in its accuracy and if the mechanism of impact or underinvestment is determined to be fairly uniform across industries, a national estimate is feasible.[16]

For the most part, however, an accurate and conservative approach for economic impact analyses of government R&D programs requires a limited focus, relating R&D expenditures to the sectors/industries that *directly* use the resulting technology infrastructure. At the levels at which R&D programs and projects are proposed, funded, and carried out, microeconomic techniques are the appropriate methodological approach for understanding key relationships such as investment incentives, lag structures of R&D expenditures (time relationships between expenditures and impacts), and eventual economic outcomes.

2.2.4. Analytical Frameworks for Retrospective Impact Assessment. Selection of a framework for economic analysis of R&D as an investment is confounded by the fact that the output of this investment does not have an explicit market (in contrast to a good or service). Moreover, the results of R&D are neither comparable across projects nor

[14] Because of the huge leverage effect of the multipliers in input-output models, small changes in variable definition, lead to large changes in results. See Griliches [1975, 1979].

[15] A panel of experts was used to judge the changes in the parameter values resulting from the R&D program for the relevant cell in the input-output matrix (transportation equipment). Using the multipliers for the other 52 cells in a REMI model of the U.S. economy, the macroeconomic impact estimates were estimated.

[16] See Section 5.2 for an example of such a microeconomic study with extrapolations to national impact estimates.

countable (Griliches 1977]). Within these constraints, selection of an analytical framework for assessing impacts of specific R&D projects frequently is determined by data availability, which results in one of two major alternative approaches being chosen:

Time Series Intervention. An ideal analytical approach is the construction of a time series of economic activity of affected industries that includes a period before government intervention. At some point in the time series, a government-funded project (R&D, technology transfer, technical information dissemination) occurs and the subsequent portion of the time series reflects the technical and economic impacts of the intervention.

The ability to effectively apply this approach depends significantly on the nature of the R&D project. Generic technologies (see Fig. 3) are typically developed early in a technology's life cycle and hence little R&D investment data are generated prior to government intervention. In fact, a major government role in most industrialized nations is to promote initial (generic) technology research through programs such as NIST's Advanced Technology Program (ATP) or Europe's Framework Program. In contrast, because certain types of infratechnologies are needed in the middle of a technology life cycle (or at least the demand for them exceeds some threshold at that point), the potential exists for obtaining data on economic activity prior to the government intervention.

However, data on economic activity "before" the intervention is frequently unattainable for either type of government project. Obviously, these data are generated farther back in time than subsequent post-intervention ("after") data. Therefore, because sources of data degenerate and eventually disappear over time, the longer the optimal time series the lower the quality of data obtainable in the "before" period (if it is obtainable at all). Thus, even when an intervention can be clearly defined in mid technology life cycle, the feasibility of collecting accurate data farther back in time than about six years is low in most technology-based industries.[17] Thus, interventions that fund new generic technologies (for example, research programs at NIH, DoE, or NIST's ATP) have no easily measurable "before", unless "defender" technologies exist at the time of the study. However, using the defender technology as the before period introduces the issue of whether its net benefits should be subtracted from the market penetration history of the new technology as it is substituted for the older one.

Counterfactual Estimation. Because availability of data and other difficulties frequently preclude the construction of a time series of economic trends before government intervention, the analyst must often use a "counterfactual" technique to estimate the differential impacts of the government R&D project.[18] In the application of such a technique, industry respondents are asked a series of "what if" questions focusing on the implications of additional costs incurred by industry if the government project did

[17] In fact, discussions with managers in some industries put a limit of three years on collections of some types of data due to the dynamic character of their industries (mergers, acquisitions, exits, labor mobility).

[18] A frequently cited early application of the counterfactual technique is Fogel's [1962] study of "social savings" from the emergence of railroads in the United States. Although much social research involves implicit counterfactuals, Fogel is recognized by economic historians as the first researcher to explicitly state a counterfactual as the basis for impact analysis.

not exist. Such a technique works well when the government project either is initiated during the current technology life cycle, so that some experience without the government contribution exists, or the project is an intervention in a life cycle that has similarities with related technologies, allowing the respondents to extrapolate from prior experience.

The counterfactual approach has been used extensively by NIST in assessing the economic impacts of its infratechnology research. Such research responds to a demand for technical infrastructure resulting from some event (such as a new technology or a regulation). Often, such needs increase in magnitude as markets expand until a threshold is passed that creates demand for the infrastructure. In some cases, a new infratechnology replaces less efficient forms used in the current or previous technology life cycles. Experience with the less efficient infrastructure being replaced or knowledge of similar infrastructure from past life cycles provides industry respondents with more accurate perspectives on the increased costs that would be incurred if the new infrastructure were not available. This approach may sound similar to the time series intervention. However, the counterfactual approach is a "second best" solution to characterizing costs in the period before interventions in situations where constructing a pre-intervention time series of net benefits from previous infrastructure (or no infrastructure) is not feasible. Usually, this is the case because annual cost data cannot be estimated or data collection is judged to be too difficult. As substitute, the counterfactual approach obtains an average annual estimate of costs in the pre-intervention period.

NIST's ATP has also used the counterfactual approach to assess the impacts of its generic technology funding on corporate R&D investment decisions. Here, the counterfactual would be no R&D project, a smaller or less ambitious project, or a time delay in funding the same project. Questions about these possible impacts typically are asked as part of a broad program impact assessment, as opposed to a single project impact study (discussed in a later section).

2.2.5. Determining the Scale and Scope of Studies. Within a particular generic analytical approach, economic impact studies can be undertaken with varied depth and coverage. Therefore, in designing a study, a number of choices have to be made with respect to scale and scope. One of the first decisions is the determination of the categories of desired information. Some studies are undertaken with a singular focus on a bottom-line quantitative impact estimate. However, while quantitative impact estimates are an important part of economic studies, semi-quantitative or purely qualitative assessments are also extremely useful. For example, information on the effects of an R&D agency's interactions with industry on such decision variables as R&D directions, market entry decisions, and product diversification strategies can provide useful information on scope and nature of impacts. Such information is useful for role justification or modification and also provides a valuable input for planning future research programs.

Another critical scope decision is the number of industries/sectors from which impact data are to be collected. Two countervailing motivations exist. If the industries studied are limited to those that receive significant and direct technical input from a government agency, the data can be collected from those industries by survey. Even if the data collected have to be extrapolated from the sample of firms surveyed to obtain an impact estimate for the entire population (an industry), at least the estimate will be based on explicit assessments by industry of the impact of the government-supplied generic

technology or infratechnology. By relying on primary data sources (directly impacted industries), such first-order impact assessments are highly credible.

Of course, the analyst will be tempted to extend the analysis to include second-order benefits that occur indirectly downstream in the supply chain from the point where the first-order impacts occur. For example, new or higher quality products (indirect impact) result from better quality components purchased from upstream supplying industries using the government-subsidized technology infrastructure and associated standards (direct impact).

Even the identification of direct beneficiaries can be complicated when a government R&D agency works with a number of industries simultaneously. The latter situation may give the impression of direct benefits having been delivered to several industries in a supply chain, but these benefits often vary with some of the industries receiving direct assistance in the development or use of the technology infrastructure, while others receive only indirect assistance such as procurement information or incentives. The key criterion for scope determination is direct vs. indirect *use* of the government-supplied infrastructure.

As discussed in a previous section, extension of coverage of an impact study can be taken to the extreme in an attempt to estimate the total (regional or national) economic benefits for multiple supply chains by use of input-output models. Such macroeconomic models assume a multiplier exists with respect to secondary benefits that accumulate through linkages among all industries. Using such multipliers to extrapolate benefits across all industries will produce large economic impact estimates because of the assumed multiplier effect. However, the small sizes of the interjections typically being studied are way below the resolution of such models and impact estimates obtained in this way are therefore not defensible.[19] Thus, economic studies should be limited to estimating the impacts on the several industries from which directly measurable impact data can be obtained.

2.2.6. Definition and Selection of Metrics. This step is critical because it drives survey design and eventual impact measure calculations. Unfortunately, it is frequently mishandled. The general approach requires decisions about the scope and heterogeneity of the technology to be studied, inclusion of subsequent categories of investment necessary to achieve commercialization, and accounting for the "cost" of scrapping the existing technology

In selecting outcome metrics for quantitative measurement of economic impacts, the ideal approach is to choose those metrics that represent the "final" impact of the diffusion of the technology infrastructure on the relevant supply chain segment. However, final or ultimate outcomes can require complex metrics, such a "quality of life index" for medical technologies and services. Such final outcome metrics represent measures of the social objectives of broad government programs, and many analysts separate government programs into those with social and those with economic objectives.

[19] Of course, such studies typically make only a crude estimate of the direct benefits, which in reality is probably sufficient given the crude and largely unjustified extrapolations to the rest of the economy.

However, this division often creates an artificial dichotomy. Economic metrics are essential for assessing impacts in both cases. Better health care is correctly labeled as a social objective, but achieving such an objective can require the creation of not just a new technology (biotechnology) but also a new industry to supply the technology. Thus, many outcome metrics for government R&D programs in support of ultimate social objectives (such as health care) will be economic. The correct framework is to regard these classes of metrics as hierarchical rather than as substitutes and make a decision as to the final metric category for the desired analysis.

Historically, government-funded R&D and subsequent government procurement in areas with social objectives such as national defense and energy independence have jump-started new industries or at least significantly expanded embryonic ones. Semiconductors, computers, and network communications are examples. The efficiency with which government R&D programs facilitate the formation of new companies and an effective industry structure will determine the efficiency with which a social objective (such as better health care) is attained. Thus, useful impact assessment in virtually all cases will require economic impact metrics.

In selecting economic impact metrics, the structure and coverage of benefits and costs is particularly important for the ultimate estimation procedure. One of the initial decisions focuses on the desirability of establishing and including a baseline of net benefits from an existing technology. For example, in studies of social rates of return from private-sector innovations, Mansfield et al [1977] argued that benefits (profits) to imitators should be added to benefits accruing to the innovating firm and that benefits lost to competitors supplying the old technology should be subtracted. Further, unsuccessful R&D by competing firms should be added to total costs.

These issues are mitigated somewhat for quasi-public goods such as infratechnologies. Infratechnologies and associated industry standards are typically introduced at various points in the technology life cycle where markets already exist. In such cases, the existing product structure is not replaced; rather, measurement of the performance of some attribute of the product or an attribute that provides an interface with other products is standardized. The resulting productivity increase can be measured as an incremental gain in an existing production process, which is, in effect, equivalent to Mansfield's requirement to net out the residual value of obsolete technology.[20]

For generic technologies, which replace existing technologies (Mansfield's focus), the issue of subtracting benefits lost requires more attention. Even here, for prospective studies, at least, a capital budgeting approach would require only estimating rates of return over the study period for both the new and defender technologies from time zero and making an investment decision accordingly (that is, R&D and other initial investments associated with the defender technology are regarded as sunk costs and ignored in the calculation). In other words, a marginal rate of return approach from the beginning of the study period is used. For retrospective studies, one also can rationalize ignoring the defender technology. What really counts is the rate of return realized by the technology under study relative to an appropriate hurdle rate.

[20] Of course, if assimilating the new infratechnology results in the purchase of new equipment, for example, writing off the old equipment could be viewed as constituting a "cost".

With respect to the requirement to include all R&D investment associated with the development of a new technology, the public good nature of an infratechnology frequently means that a single government-funded or government-industry cost-shared project is undertaken to produce the required infratechnology. That is, the equivalent of "unsuccessful competitor" R&D does not exist, unless several companies were funding the infratechnology R&D in the absence of or in addition to a government or industry-wide effort. If the latter is the case, that funding should be added to total social costs. Any "pull" costs required to assimilate the resulting infratechnology/standard should be estimated and included.

Similarly, generic technology research may be carried on simultaneously in individual firms and supplemented by government research undertaken independently, or government funding programs may subsidize only some of the individual research efforts within the industry. In such cases, the entire generic technology research expenditure should be included, if the objective is to estimate the impact of the generic technology on subsequent economic performance. When the objective is to estimate the impact of the government role, then two calculations should be made—one for the total impact and one for the contribution of the government subsidy. The former is analogous to the Mansfield framework, while the latter examines the relationship between government costs (or government-industry costs, if the research project is a collaboration) and the portion of the resulting economic benefits attributable to that project.

In general, the selection of specific metrics for an economic impact study is determined by the following factors:

- objectives of the study
- nature of benefits and costs
- available assessment expertise
- resources provided for the study
- quality of primary data sources (both benefits and costs)

Alternative classes of metrics differ in qualitative and quantitative content. As a general strategy for quantitative analysis, one of three levels can be selected. At the most basic level, descriptive statistics such as additional amount invested in R&D in response to the government project, acceleration in R&D investment, number of patents or new products introduced, percent increase in productivity, or percent reduction in market transactions costs. If the quality of the data permits, a second level can take some of these statistics and calculate measures such as those used in corporate finance. These measures of impact offer a basis for comparison to some degree across projects. Finally, for high-quality time series, formal mathematical models of the impact (based on, say, a production function) can be used. This last approach offers precise estimations of the contributions of specific inputs taking into account interactions with other inputs.

For a particular category of metrics (inputs, outputs, and outcomes), the quality of a specific selection will be determined by the following criteria:[21]

(1) *Simple and Computable:* analysts can learn to apply the metric in a straightforward and easy manner.
(2) *Persuasive:* the metric appears to be measuring the correct attribute. In other words, it displays face validity.
(3) *Consistent and complementary:* the results are reproducible and complement other metrics.
(4) *Technology and industry independent:* the metric should not be based on specific technologies or stages of economic activity where the intervention occurs but should be based instead on consistent performance attributes that characterize industrial strategies and performance in general.
(5) *Gives feedback:* results from the metric provide useful information to the analyst performing the test and to the relevant stakeholders (program managers, policy makers).

Very important is the classification of metrics by stage of economic activity. This taxonomy is necessary for selecting compatible metrics across these stages. Following the GPRA classification scheme, the metrics required for a complete impact study fall into three categories:

Input (cost) metrics. All costs, private and public, should be included. Some cost data may have to be disaggregated and apportioned to the project under study and other projects. Specific cost categories are

- direct and indirect government research program costs: research labor, production labor (for prototypes and other transfer artifacts, such as standards), overhead, equipment, and technology transfer/outreach
- industry research program costs: research labor and overhead (for independent or joint research projects), "pull" (technology assimilation) costs, including fees paid to government for technology transfer artifacts and related services
- industry commercialization costs: applied R&D investments, capital costs, workforce training costs,

Output metrics. Conducting economic impact studies of government research requires the selection of performance variables that can be directly attributed to the government funded/conducted research project and that can be related to subsequent economic impacts (outcomes). Examples of output measures frequently identified are

- contributions to underlying science
- generic technology or infratechnologies developed
- percent of companies in target industries assimilating/using generic technology or infratechnology

[21] Derived from Pressman [1992].

- measures of intellectual property and its dissemination resulting from the research project, such as patents or licenses in the case of generic technology and adoption of standards in the case of infratechnologies
- promulgation of industry standards

Outcome metrics. Selection of specific outcome metrics depends on a number of factors, including the type of R&D targeted by the project being studied (in particular, generic technology vs. infratechnology) and the objectives of the broader research program of which the project is a part (which may include industry structure and growth objectives). Categories of outcome metrics frequently estimated include impacts on

- industry R&D investment decisions
- market access and hence market entry decisions
- industry cycle times (time to market)
- productivity (R&D or production process)
- rate of market penetration of new technology
- product quality
- product and system reliability
- transaction costs (equity in trade, performance verification)

For complete impact assessments, a set of *measures* must then be selected that both summarize the economic impacts derived from the metrics and allow comparisons with a reference standard for minimum efficiency.[22] Such measures vary in terms of their quantitative content, ease of calculation, and ultimately type and quality of information provided. Classes of outcome measures in order of increasing quantitative content and explanatory power include

- **peer review assessments** of the nature and relative magnitude of impact—yield largely qualitative information and, at most, ordinal rankings across projects
- **customer satisfaction surveys**—yield largely qualitative (normative) impact assessments
- **corporate finance measures** (net present value, benefit-cost ratio, internal rate of return)—yield quantitative measures in the form of averages across respondents' companies (and, in some cases for respondents' customers/suppliers)
- **microeconomic models** (productivity, sales, profits, employment, value added)—yield quantitative estimates that can be related to specific outputs of the research project and to the collective set of inputs (usually industry level, but company characteristics can be included in the model estimated)

2.2.7. Estimation of Metrics (Calculation of Measures). Microeconomic modeling is the ideal approach to estimating economic impact measures because the relationships

[22] Metrics are variables that represent impact. Measures are calculations that summarize the relationship between cost metrics and benefit metrics.

among inputs, outputs, and outcomes are explicitly specified. However, the types and minimum quality of data required for meaningful impact estimation are greater than those needed for corporate finance measures. The latter require fewer and somewhat less precise individual (annual) values of inputs and outcomes and allow more liberal extrapolation of available data to achieve desired time series.[23] Further, the fact that interactions are involved in the model's structure means that data on all inputs must be of equal quality. These requirements are difficult to meet. Another reason favoring corporate finance metrics is the fact that corporations use these measures, thereby providing some basis for comparisons of the results of industry R&D investments with supporting government projects. Also, industry understanding and interpretation of impact study results are facilitated, thereby enhancing industry's support for future R&D programs. For these reasons, the NIST economic impact studies chose to use this category of metrics.

If collected data permit, three corporate finance measures should be calculated in each study: net present value (NPV), benefit-cost ratio (BCR), and internal rate of return (IRR). All three measures can be derived from the same time series of benefits and costs, but calculating all three is rationalized by the different perspectives provided on impact magnitudes and patterns over time.

Brief definitions of the three measures are given below. They are discussed in more detail with examples in Section 2.3.

Net Present Value (NPV): The NPV, or discounted cash flow (DCF) as it is often called in industry, is the value of the inflation-adjusted (real) net benefits produced by a project over a specified time discounted to the current (present) year or to some other reference year. NPV provides an absolute value for the economic benefits produced and therefore is theoretically the most accurate measure of economic value. It is also relatively simple to calculate and interpret. The time preference of money is incorporated in the calculation by selecting a discount rate (also referred to as the "opportunity cost of capital"). The selection of the discount rate is therefore a critical decision for the analyst.

Benefit-Cost Ratio (BCR): The BCR is the ratio of the present value of real benefits to the present value of real costs. It allows comparisons among projects and is the preferred metric when assessing ongoing projects without well-defined beginning and termination points. Its disadvantage to some users is that it does not provide a sense of yield per time period (and hence cannot be directly compared to other "rates," such as a minimum acceptable rate of return or hurdle rate). The BCR also depends on a reasonably "correct" selection of a discount rate (as does NPV). Finally, it is quite complex to use for project selection compared with the NPV.

Internal Rate of Return (IRR): The IRR is the discount rate that makes the NPV of a project zero (the equivalent of benefit-cost ratio of one). It provides a measure of the yield from an investment per time period and allows comparisons

[23] Primarily for these reasons, corporate finance measures have been used extensively in the NIST impact studies.

with other yields or rates such as the opportunity cost of capital. A big advantage is that it does not require the specification of an external discount rate, as do the other two measures. It depends only on the internal characteristics of the project being analyzed. However, it can malfunction when used as a project selection device, especially when comparing projects with dissimilar expected benefit patterns or when the benefits from a particular project display a change in the direction of the rate of growth (i.e., an inflection point) over the planning period. Moreover, the calculation of the IRR is biased in favor of projects (such as process innovations and some infratechnologies and standards) where the net benefits are attained relatively quickly compared to those from product innovations. Even when little or no additional benefits are realized in succeeding time intervals, front-loaded net benefit time series cause the calculated value for this measure remains near its peak value. As a result, product innovations and more radical technology investments of all types, which typically realize larger net benefits but do so more slowly, can be discriminated against by this measure.

For assessing the economic impacts of government R&D projects, the discount rate used for NPV and BCR calculations is usually specified by OMB. The IRR is the only metric that estimates a discount rate (hence, the name "internal" rate of return) and is therefore not dependent on assumptions about external (to the project) market interest rates.

All three measures can be used to represent quantitative economic impacts, at least at the project level. Each of these impact measures has strengths and weaknesses, which is why all three should be used together. For example, NPV is an absolute value referenced to a specified year (to adjust for inflation and the time preference of money). It therefore provides a straightforward and unambiguous perspective on the relative magnitude of the economic impacts for approximately *equal* investments in different categories of technical infrastructure, or for unequal investments if used in capital budgeting to select among mutually exclusive projects. The BCR and IRR are efficiency measures and thereby provide different insights for project selection or retrospective assessment. Several versions of NPV exist and several distinctly different alternatives to the IRR are used by some analysts.[24] Section 2.3 provides more detail on these issues.

2.2.8. Integration of Metrics into the Analytical Framework. As previously stated, microeconomic analyses of the technology, industry structure, and competitive dynamics are essential in order to first select the appropriate metrics and then to place them in a context that allows a complete understanding of their roles and impacts. Fig. 4 provides an example of the range of metrics that can be used to describe and estimate economic impact over a technology's life cycle. Some of the metrics are specific to and hence measured directly from the project, such as inventive and innovative output, return on

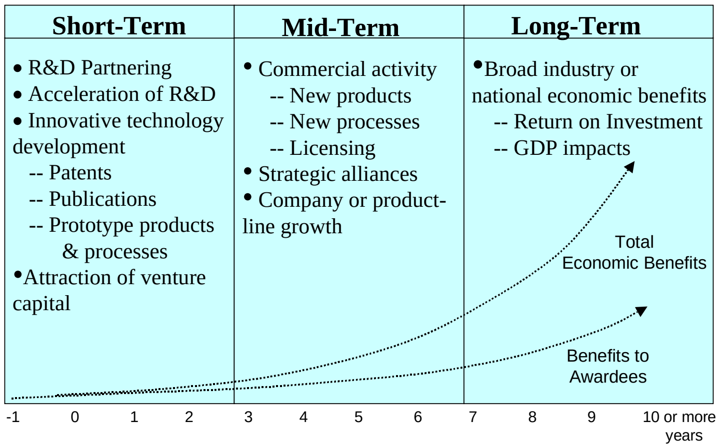
[24] In economic impact analysis, the IRR is called the private rate of return (PRR) when applied to a single firm and the social rate of return (SRR) when applied to the innovator, subsequent imitators, and the users of the technology (i.e., one or more industries or even the entire economy). The difference between the two in terms of economic benefits realized is a rough measure of the degree of diffusion of the technology beyond the innovator. The SRR is usually the main metric for government research programs because the impact target is at least the industry level.

investment, or GDP impacts. Others, such as strategic alliances or venture capital availability, are contextual in that they describe institutional interfaces or link the role/impact of the project to other investments and infrastructure in the relevant industries.

Figure 4

Organization of Metrics by Technology Life Cycle:

NIST's Advanced Technology Program



Source: Adapted from Ruegg [1999, p. 19]

2.2.9. Program vs. Project Metrics. In depth and quantitative analysis is conducted at the project level. However, periodically, it is necessary to assess the program of which projects are the primary implementation mechanism. In other words, some aggregation of project analyses is required to provide a broader (program-level) assessment that matches the economic rationale analysis that created the program as a public policy instrument.

In many instances, the cumulative effect of multiple project impact studies is deemed sufficient to judge the overall effectiveness of the program.[25] However, complex or controversial programs often need specific program-level indicators, as well. NIST's ATP is an example of a program with a complex economic role rationale that is applied to a wide scope of technologies and industries. Thus, ATP's program impact assessment must be multi-layered. To this end, the Program conducts project impact assessments that use some of the quantitative tools described in this report. These analyses are often applied to a number of projects in a single technology focus program. The projects can extend over several phases of the R&D cycle, as indicated in Fig. 4.

ATP also periodically surveys both funded firms and unsuccessful applicants to obtain data for descriptive statistics that help characterize the role and impact of the program as a whole. Such metrics include R&D investment-related questions like impact of the funding on yes-no investment decisions, the timing of the investment decision, leveraging effects on subsequent private financing, etc.

2.3. Extended Discussion of Metrics

Much debate and confusion exist over the appropriate use and interpretation of the basic corporate finance metrics, so this section assesses each of them in more detail.

2.3.1. Net Present Value (NPV). The NPV of an investment is a simple criterion for deciding whether or not to undertake an investment. NPV answers the question of how much cash an investor would need to have today as a substitute for making the investment. If the net present value is positive, the investment is worth taking on because

[25] The degrees to which a program is complex and controversial are often correlated.

doing so is essentially the same as receiving a cash payment equal to the net present value. If the net present value is negative, making the investment today is equivalent to giving up some cash today and the investment should be rejected. If the projected return on an investment is identical to the selected *discount* rate, the NPV=0 and the investor is indifferent with respect to making the investment.[26] This case of NPV=0 is used as a reference point for all major metrics, including the IRR. Retrospectively, the same criteria are used to decide if an investment "paid off".[27]

In the simplest case, the present value of a future cash flow is the value of that cash flow after considering the appropriate market interest (discount) rate:

$$PV = \frac{C_1}{1+r} \qquad [1]$$

where C_1 is cash flow at date 1 and r is the discount rate. The net present value of an investment is the present value of the investment's future cash flows, minus the initial cost of the investment:

$$NPV = -Cost + PV \qquad [2]$$

The term *net present value* emphasizes that fact that the cost of the investment has been taken into account in determining its value. Thus, it is not simply the cash flow generated. Calculating the NPV requires the selection of an interest rate to adjust cash flows accruing in different years for the time preference of money. For example, if an investment of $30,000 today will produce a total cash return of $40,000 in one year's time, $40,000 is the *future value* of the investment. In contrast, the present value of the $40,000, if the market interest rate chosen as the discount rate is 10 percent, becomes

$$\$40{,}000/1.1 = \$36{,}364$$

The market rate of interest adjusts or discounts the future value of an investment. This rate is usually the rate on a riskless asset, such as government debt with a maturity of approximately the same length as the investment (R&D project) under study. The return on the riskless investment represents the *minimum acceptable alternative* to the investment under study. Thus, it is the implied rate of return that the prospective investment (or, retrospectively, the investment being evaluated) must exceed to be considered worthwhile (or, retrospectively, a success).

The calculation that reveals which situation is the case is the *net present value (NPV)*. In the above example, the NPV of the investment is $36,364 minus the original investment, so that

$$Net\ \ present\ value = \$36{,}364 - \$30{,}000 = \$6{,}364$$

[26] Corporate finance texts provide extended discussions with examples. See Ross, Westerfield, and Jaffe [2002, pp. 56–62].

[27] Note that the discount rate adjusts only for the time preference for money. It does not adjust for inflation, which is accomplished by an appropriate inflation index.

Many investments, including R&D projects, generate both costs and cash flow over time. Therefore, *NPV* is expressed more generally as a time series of *T* net benefits, NB_i, discounted by a selected market rate of interest, *r*.[28] Some formulas for NPV show an initial investment, NB_0, that occurs at the beginning of the project period. This could, for example, reflect investment in research facilities. However, because these facilities are typically used for many research projects over a long period of time, such an expenditure is often regarded as a sunk cost with respect to individual projects being evaluated and is therefore not included in the *NPV* calculation. Thus, NPV is frequently represented by

$$NPV = \frac{NB_1}{1+r} + \frac{NB_2}{(1+r)^2} + \frac{NB_3}{(1+r)^3} ... + \frac{NB_T}{(1+r)^T} = \sum_{i=1}^{T} \frac{NB_i}{(1+r)^i} \qquad [3]$$

In finance, such a time series can be a *perpetuity*, which is a constant stream of cash flows (or net benefits) without end. Such a *geometric* series has an infinite number of terms, although the whole series has a finite sum because each term is only a fraction of the preceding term. In evaluating the economic impact of an R&D project, this situation would be an unrealistic case, even though an argument could be made that, once created, knowledge never loses its utility. However, from a practical economic impact perspective, the technology created by an R&D project has a finite life cycle.

This is because, while the discovery of *knowledge* remains in perpetuity, *technology* depreciates as economic conditions change, creating a demand for new technology.[29] When the new technology arrives, the defender technology is declared obsolete and is no longer used in an economic sense (even though the knowledge it embodies remains). The analyst therefore attempts to define the *economic* life cycle for the technology or technology element under study and uses that life cycle or a truncated portion of that cycle as the time period in the impact assessment.

Whatever study period is selected, the pattern of net benefits differentially affects the basic choices among measures of economic impact and thereby confuses comparisons among projects. Even for one metric, such as NPV, selection of the study period is critical. For example, in corporate finance, the payback period rule is often used to constrain investments to those that "payback" in a specific period of time. A time period is arbitrarily selected in which a proposed investment must generate a positive NPV to be approved, and thus projects that produce substantial cash flows relatively late in feasible planning cycles (particularly true for R&D) could be rejected.

2.3.2. Benefit-Cost Ratio (BCR). The benefit-cost ratio takes the times series data on benefits and costs used to construct NPV and organizes them in ratio form rather than as an absolute value. Alternatively, the BCR can be defined as the ratio of the discounted benefits to the discounted costs of an investment with reference to the same point in time. As a result, the BCR can provide an indication of relative efficiency among similar projects. This metric also requires the explicit selection of a discount rate. The general definition of the BCR is represented by Equ. [4].

[28] For almost every conceivable R&D project, *r* is an annual percentage rate and the both benefits and costs are considered to occur at a single point in time for each year in the time series.

[29] Tassey [1997, p. 68].

$$[4] \qquad BCR = \frac{\sum_{i=1}^{T} \frac{B_i}{(1+r)^i}}{\sum_{i=1}^{T} \frac{C_i}{(1+r)^i}}$$

Like the NPV method, the BCR cannot be readily compared to an opportunity cost, which is typically stated as an annual percentage rate. It is an inferior measure to the NPV and IRR methods because of its sensitivity to relatively small variations in the aggregated benefit and cost time series and therefore requires complex calculations in the project selection mode. This sensitivity is the result of the fact that the BCR is computed as a single ratio. An outlier, an error in measurement in a single year, or mislabeling benefits and costs can significantly affect the numerator or denominator and therefore the value of the BCR itself.

Mislabeling costs as benefits and vice versa might seem unlikely, but it happens. Such a situation occurs when a portion of the estimated benefits from a project can be characterized alternatively as a benefit or a reduced cost and therefore either added to the numerator or subtracted from the denominator, respectively.[30] This situation has occurred in two NIST economic impact studies of projects where both industry and government contribute portions of the cost of a research project, but the industry funds are labeled and collected as fees, rather than as direct contributions to the research. Such fees could be considered a cost (added to the denominator) or a negative benefit (subtracted from the numerator). The two alternative approaches produce decidedly different BCRs.

The problem can be seen clearly in the following example. Suppose the benefits profile has been established except for one item. Let the present value of this item be *d*, which may be interpreted either as an additional benefit or a reduced cost. The present values of the benefits and costs of the remaining items in the cash flow profile are denoted by b and c, respectively. Then, consider the following two cases: (1) *d* is an additional benefit; (2) *d* is a reduced cost. The NPV in the two cases is the same:

(1) $NPV = (b+d) - c = b - c + d$

(2) $NPV = b - (c-d) = b - c + d$

On the other hand, computing the BCR for the two cases yields different ratios:

(1) $BCR = \frac{b+d}{c}$

(2) $BCR = \frac{b}{c-d}$

This problem can be important in a retrospective economic impact assessment. However, such classification problems may or may not affect the use of the BCR for prospective assessments of alternative (independent) research projects. If

[30] Au and Au [1992, pp. 191–192].

$$\frac{b+d}{c} \geq 1 \qquad (i.e., b+d \geq c)$$

then

$$\frac{b}{c-d} \geq 1 \qquad (i.e., b \geq c-d)$$

That is, if a project is acceptable on the basis of the first ratio, it is also acceptable on the basis of the second ratio. Therefore, the economic feasibility of the project will not be altered, although its ranking certainly could be changed.

2.3.3. Internal Rate of Return (IRR). In corporate finance, the *internal rate of return* (IRR) is considered the most important alternative to net present value. For purposes of analyzing the economic impacts of R&D, the IRR is alternatively called the private rate of return (PRR) when the return to a single company's (the innovator's) R&D investment is being studied, or the social rate of return (SRR) when industry-wide or economy-wide rates of return are estimated. The latter is typically of greater interest to R&D agencies because such estimates capture the economic growth impacts of their projects, provide a potential basis for R&D project selection, and reflect on an agency's mission.

The basic rationale behind the IRR is an attempt to construct a single metric that summarizes the merits of a project. Unlike other metrics, this number does not depend on the interest rate that prevails in the capital market. This is why it is called the *internal* rate of return; the number is internal or intrinsic to the project and does not depend on anything except the cash flows of the project.

Technically speaking, the IRR is an absolute measure of the percentage rate of net benefits from an R&D project over the project's lifetime.[31] A simple example would be a project in which \$100 is invested in time period 1 and creates a value of \$110 in period 2. Using Equ. [2] yields

$$NPV = -\$100 + \frac{\$110}{(1+r)}$$

where r is the discount rate. The IRR is the value of r that makes the NPV of the project equal to zero. Choosing NPV=0 as the reference point for calculating the IRR is not arbitrary because it represents the breakeven condition for the investment.[32]

The process of calculating the IRR involves trying different discount rates until the rate is found that equates NPV to zero. Using the above example, substitution of an arbitrary discount rate of 0.08 gives

$$NPV = -\$100 + \frac{\$110}{1.08} = \$1.85$$

[31] For an extended discussion, see Au and Au [1992, Chap. 8].

[32] This equivalent to a BCR=1.

Since the NPV is positive, one would next try a higher discount rate, say 0.12. Using this rate yields

$$NPV = -\$100 + \frac{\$110}{1.12} = -\$1.79$$

The trial and error procedure eventually tells us that the NPV of the project is zero when r equals 10 percent.[33]

The general investment rule is

> *Accept the project if the IRR is greater than the discount rate. Reject the project if the IRR is less than the discount rate.*[34]

The above example is a simple one-period project, which generates only one cash flow payment. For most investments, the resulting cash flow occurs over several periods. In such cases, the IRR is the unknown in the following equation:

$$NPV = 0 = \frac{NB_1}{1+IRR} + \frac{NB_2}{(1+IRR)^2} + \ldots \frac{NB_T}{(1+IRR)^T} = \sum_{i=1}^{T} \frac{NB_i}{(1+IRR)^i} \qquad [5]$$

Considerable confusion exists as to the precise meaning of an IRR. Although it appears as an annual discount rate in the formula, its correct interpretation is as a percentage yield occurring over a defined period of time. Thus, unlike NPV, which is an absolute value, the IRR is stated as a percentage and therefore implies an efficiency measure of invested funds over time. Alternative projects can be compared against a discount or hurdle rate and, under certain conditions, against each other for a given planning or study period.

The IRR is *not* comparable to a compound rate of interest, as is sometimes assumed. This confusion arises because discounting is the opposite of compounded growth, and it is therefore often erroneously assumed that the IRR is the compound annual yield over the evaluation period. In an investment that compounds, the original amount (principal) is reinvested in every succeeding time period along with the accumulated interest, as in a conventional bank account. Thus, with compounding, a relatively small interest rate can generate a large cash balance over time.

In calculating an IRR, net benefits are in effect withdrawn from the "project account" in each period as they are generated. These funds are assumed to be reinvested at some external (to the project) rate of return, which is the opportunity cost of capital. Only the

---

[33] One could directly solve for r in this example after setting NPV equal to zero. However, with a longer series of cash flows, it is not possible to directly solve for r. Instead, a trial-and-error approach similar to this example must be used. See Ross, Westerfield, and Jaffe [2002, pp. 147–149].

[34] For corporate R&D projects, the discount rate is typically not the market rate of interest (also referred to as the "cost of capital"). Companies establish "hurdle" rates that reflect past performance from the same class of investment and/or what rate or return investors demand to adjust for risk. That is, the market rate of interest does not represent the minimally acceptable rate of return on these funds and is therefore not an appropriate hurdle rate. For government R&D projects, the opportunity cost of capital is the OMB-specified discount rate.

funds that remain invested in the project at the end of each period will earn returns in the following period at the IRR.

In the case of an R&D investment by a company or a government agency, an initial investment is made (even if it occurs over several years) and the resulting cash flow goes into a separate account (the corporate account in the case of the PRR and the national economy account in the case of the SRR). That is, the "principal" (the original investment) is not reinvested in the project. In fact, this investment is often a "sunk cost", with little or no value at the end of the investment period. Thus, it is incorrect to apply an IRR realized in one project to estimate the "returns" from possible future projects, as if it were a compound rate of interest.

Several additional problems can occur in applying the IRR. The most commonly cited problem is the possible existence of multiple solutions. The first few NB_i in Equ. [5] can be negative, reflecting the dominance of investment costs. However, once cash flow turns positive, it must remain so to avoid multiple values for the estimated IRR.[35]

Another complaint about the IRR has to do with the relative scale of different investments being studied simultaneously. Consider the two alternative investments in Table 1. Project 1 costs $10 million and Project 2 $25 million. Assuming a 25 discount rate to represent the relatively high risk of R&D investment, Project 1 yields an NPV of $22 million and Project 2 yields $27 million.

Table 1

Scale Effects in Evaluating R&D Projects

	Cash Flow at Beginning of Project	**Cash Flow at End of Project (one time period)**	**NPV @ 25%**	**IRR**
Project 1	-$10M	+$40M	$22M	300%
Project 2	-$25M	+$65M	$27M	160%

An important consideration is whether the two projects are "independent" (acceptance or rejection of one project is independent of acceptance or rejection of other projects) or "mutually exclusive" (only one of the projects, at most, can be accepted). The latter could be the case if, say, only one laboratory is available and both projects require that facility. Assuming the two projects are "mutually exclusive", corporate finance texts say that Project 2 is the superior investment because the general rule is to rate projects by NPV.[36]

An apparent problem with this decision is the fact that Project 1 has a much higher IRR. However, an important insight into the selection of Project 2 can be gained by

[35] From the theory of polynomials, the *n*th order polynomial has *n* roots. Each such root that is positive and less than 1 can have an IRR associated with it. Applying Descartes's rules of signs gives the result that a stream of n cash flows can up to *M* positive IRRs, where *M* is the number of changes of sign for the cash flows (Ross, Westerfield, and Jaffe [2002, p. 152]).

[36] Ross, Westerfield, and Jaffe [2002, pp. 149–155].

calculating the *incremental* IRR for the additional cost of Project 2 compared to Project 1. That is,

$$0 = -\$15M + \frac{\$25M}{1 + IRR}$$

From this equation, the incremental IRR is determined to be 66.7 percent, and the NPV for the incremental investment is

$$NPV = -\$15M + \frac{\$25M}{1.25} = \$5M$$

Project 1 is acceptable as an *independent* project because its NPV is positive. However, as mutually exclusive projects, investing the additional $15M to fund Project 2 is the correct decision because the NPV is positive, and the incremental IRR is higher than the discount (hurdle) rate of 25 percent. In summary, Project 2 is the superior project based on any of three criteria: its NPV is larger, the incremental NPV is positive, and the incremental IRR is greater than the discount rate.

2.4. Use of Economic Impact Measures to Assess Government R&D Programs

The above methodological issues faced by policy makers in managing economic impact assessment programs can be made clearer by an example comparing different types of government R&D projects. Table 2 provides such a comparison. The three projects presented support R&D targeted at different elements of the typical industrial technology. Project (1) supports generic product technology development, Project (2) supports generic process technology development, and Project (3) supports infratechnology research. Each type of research typically produces a unique pattern of net benefits over the technology life cycle.

Unfortunately, interpretation of conventional impact metrics is made difficult by the different behavior of commonly used measures over the study period.[37] In Table 2, Project (1) supports eventual product innovation. Product innovations can yield large net benefits, but they typically occur relatively slowly due to the process by which innovations penetrate markets (by displacing existing or defender technologies and overcoming typical customer uncertainty with respect to the new technology's performance). Project (2) supports process innovation. Benefits are typically realized sooner because process technologies are frequently introduced in the middle of the technology life cycle when volume is expanding and quality control and cost reduction

[37] A complete technology life cycle is not represented. The ten years covers the development, commercialization, and market penetration up to the point of market saturation (indicated by a leveling of net benefits. In subsequent years, net benefits will decline as the technology becomes obsolete and is eventually replaced.

Table 2

Net Economic Benefits from Three R&D Projects ($millions)

Projects	Net Benefits by Year from Start of Project 1	2	3	4	5	6	7	8	9	10	NPV @7%	NPV @ 25%	NPV @ 50%	IRR 5 yrs	IRR 7 yrs	IRR 10 yrs
(1) Product Innovation	-10	-10	-10	10	80	200	320	400	425	425	1,051	300	69	44%	100%	112%
(2) Process Innovation	-5	-5	4	20	40	50	80	80	80	80	252	79	21	85%	109%	115%
(3) Infratechnology	-2	-1	3	8	12	18	23	23	23	23	78	25	7	100%	119%	123%

are increasingly important competitive strategies, thereby creating imperatives for such technology. This pattern is indicated in Table 2 by a more rapid escalation of positive net benefits for Project (2).

Project (3) develops an infratechnology that becomes part of the industry's (or several industries') technological infrastructure. Infratechnology research projects can often deliver elements of the eventual complete infrastructure quite rapidly (for example, an initial test method or database), with other elements or improvements on the initial elements provided over the technology's life cycle. Thus, infratechnologies often can deliver some positive net benefits relatively quickly, with a build up to a steady-state level by mid cycle.

The different patterns of net benefits exhibited by the three research projects affect the alternative measures of economic impact differently. In Table 2, three estimates of net present value (NPV) are presented for three different discount rates. The 7 percent rate is mandated by the Office of Management and Budget (OMB) for assessing the economic impact of government programs. A 25 percent discount rate is used to approximate a hurdle rate for corporate R&D projects and for venture capitalists. Finally, 50 percent is chosen as a conservative social hurdle.[38] The effect of the selection of the discount rate on the NPV estimates is evident. However, for even the highest discount rate, all three projects are justified (positive NPV). Note that the scales of these projects are quite different and Project (1) would be selected for the mutually exclusive case.

The examples in Table 2 also demonstrate one of the problems with the IRR, namely, its sensitivity to the length of the planning or impact assessment period. Even though the product innovation generates the largest NPV, it would not meet the SSR hurdle rate if a five-year payback period were used. Similarly, if the research project were evaluated 5 years after its inception, it could be judged a failure based on this criterion. However, if the planning/impact assessment period were 7 years, the project is easily justifiable. In contrast, the process innovation and infratechnology projects are justified in a five-year planning/evaluation period, even though their NPVs are smaller and their long-term IRRs (10 years) converge with that of the product innovation. This happens because the process innovation and infratechnology projects generate positive net benefits sooner.

The effect of the planning period can obviously have a significant effect on corporate decisions with respect to the type of R&D projects undertaken. Longer-term, higher-risk R&D can result in product innovations that have significantly greater payoff. However, if the corporate planning horizon is sufficiently short, such projects will frequently not be funded because the major economic benefits are realized too far in the future, given the discount rate applied. In other words, the combination of a high risk premium and a bias in the IRR calculation towards early net benefits reduces the rate of return estimate below the hurdle rate. In summary, the length of the planning/evaluation period strongly affects the IRR calculation, as does the pattern of the net benefits over that period.

Moreover, for certain realistic the patterns of net benefits, the NPV and IRR will rank projects of equal cost differently. Consider the two research projects in Table 3, one for a product innovation and the other for a process innovation applicable to the current

[38] Choice of a social discount rate is discussed in more detail in a later section.

generation of product technology. These two projects are not complementary and they are mutually exclusive in that the R&D budget can fund only one of them. If they are evaluated based on the OMB-imposed opportunity cost of capital, the NPV estimate is greater for the product innovation project, but the IRR is substantially greater for the process innovation project.[39] In contrast, if the private sector opportunity cost of capital for R&D investments of 25 percent is used, the NPV estimate is now higher for the process innovation project and agrees with the IRR estimate that this project is superior.

Table 3
Reversal of Rankings of R&D Projects

Projects	Net Benefits by Year ($millions)						NPV @7%	IRR 6 yrs	NPV @25 %
	1	2	3	4	5	6			
(1) Product Innovation	-10	0	10	20	45	120	126	108%	52
(2) Process Innovation	-10	25	35	35	35	35	116	276%	61

Each impact measure is sensitive in different ways to the pattern of net benefits estimated for the study period. The different effects on the three basic measures from a change in a time series of net benefits were evident in one of the NIST retrospective economic impact studies in which the annual costs of the government program were revised to include omitted industry cost elements. The originally included research costs were incurred in years 1 through 3 and the three measures estimated. Additional costs in the form of industry certification costs incurred in years 4 through 16 were added later and the metrics recalculated. The SRR and NPV changed a little but the BCR declined significantly, as shown in Table 4.

Table 4
Differential Impacts on Metrics

	NPV	BCR	IRR
Estimated Metrics with government research costs	$65.3M	411:1	235%
Estimated Metrics with research costs plus industry certification costs	$63.1M	27:1	228%

These differential impacts on the three impact measures occurred because the SRR is typically influenced to a significantly greater degree by the first few years of the time series (the benefit and cost estimates for the first three years did not change). Thus, if large gross benefits occur in the first few years of the study period, the SRR calculation may not be affected much by the cost revisions, which are distributed more over time. The BCR, on the on the hand, because it is a ratio is affected more by the change in total

[39] The OMB specified nominal discount rate covering the length of time required for assessing most R&D projects was 7 percent for the period 1992–2002.

costs (although it still makes some difference where in the time series the additional costs occur due to the effect of discounting). Because NPV is a polynomial rather than a single ratio, it should be less affected than the BCR by additional costs in some years of the time series.

2.5. Selection of the Discount Rate

It is assumed, although not explicitly represented in the formula for calculating the IRR, that the estimated percentage yield on an investment for the defined time period must exceed the “opportunity cost of capital”. This opportunity cost is the *minimum acceptable* rate of return (MARR) obtainable from the set of available investment options. In corporate finance, the MARR is sometimes referred to as the “hurdle rate”, which is a MARR for similar types of investments (in this case R&D). This rate reflects the character of the particular class of investment, including a risk premium. Such a definition of the MARR excludes other possible uses of corporate funds.

For government investment projects, the MARR is officially designated by the Office of Management and Budget (OMB) to be the market rate of interest which the government pays to the public (and thus takes away from other "opportunities") in order to raise the necessary funds for the investment project.[40] Expected economic benefits are therefore “discounted” using this rate for the opportunity cost of capital or MARR. If the estimated rate of return turns out to be just equal to the “true” discount rate, i.e., the opportunity cost of capital, then decision makers will be indifferent with respect to conducting the project.[41]

However, use of a hurdle rate (localized MARR) that is appropriate for a particular category of investment seems to make more sense. It is true that any government investment that exceeds the OMB discount rate *ex poste* returns a greater benefit than the cost to the government of those funds. However, each category of investment faces a different risk profile and therefore *ex ante* should require a different expected rate of return (i.e., a hurdle rate that represents the unique characteristics of the relevant class of investment). Venture capitalists, for example, demand a higher expected IRR (25-50 percent) than do corporate R&D managers (about 20-25 percent) because venture capital is concentrated in fewer and usually higher risk R&D investments (start-up firms pursuing single, high-risk technology development and commercialization projects). Because economic research indicates that the average SRR for R&D is approximately 50

[40] OMB Circular No. A-94 provides discount rates for different evaluation time periods. These rates are basically the interest rates paid by the Treasury on its debt with a maturity comparable to the expected life of the project. Thus, if the economic benefits from a research project were to be realized over a 10-year period, the appropriate discount rate would be the 10-year Treasury note rate. However, this circular is only revised occasionally, the last two times being 1992 and 2002. The Treasury rate represents a riskless rate of return and therefore is arguably the *lowest* acceptable yield among alternatives for the available funds.

[41] In reality, a single, constant rate of discount to represent the time preference of money is quite crude. Economists have researched the issue of an appropriate discount rate for decades but are far from a consensus on what would constitute an improvement over current practice. For an exhaustive review and assessment of this literature see Frederick, Loewenstein, and O’Donoghue [2002].

percent (Nadiri [1993]), that number seems like a logical hurdle rate for government research projects.[42]

2.6. Use and Selection of Impact Measures

The previous sections have discussed alternative economic impact measures. Such measures are appropriate for R&D projects whether they are private, public, or collaborative efforts between industry and government.

Table 5

Percent of CFOs Using a Specific Impact Metric[43]

Measure	Percent Used Always or Almost Always
Internal rate of return (IRR)	75.6%
Net present value (NPV)	74.9
Payback period	56.7
Discounted payback period	29.5
Accounting rate of return	30.3
Profitability index	11.9

However, systematic quantitative assessment of the economic impacts of government R&D is just emerging. Corporations, on the other hand, regularly apply quantitative metrics to prospective and retrospective assessment of their R&D portfolios. Table 5 summarizes the results of a survey of chief financial officers (CFOs) to determine the frequency with which different metrics are applied. The IRR and NPV are dominant, being used by three quarters of CFOs surveyed. In fact, studies show that the two are often used together.

The other metrics in Table 5 are used to a lesser extent for good reasons. The payback period requires that the original investment be returned in nominal dollars in a specified number of time periods. The metric makes no adjustment (through discounting) for the order in which net benefits are delivered. The discounted payback period uses discounting for the ordering of net benefits, but it still cuts off the net benefit time series at an early point (payback) in the technology's life cycle, thereby frequently rejecting investments (especially product innovations) that exhibit a slower acceleration of net benefits but eventually produce a greater payoffs.

[42] However, as pointed our by Griliches [1995], the methodologies and quality of data used by economists in rate-of-return analyses have varied significantly. The two studies estimating internal rates of return from R&D across a wide range of industries (Mansfield et al [1977] and Tewksbury et al [1980]) found average SRRs in the 75-100 percent range. SRRs estimated in the last 10 NIST studies of infratechnologies (conducted from 1998 to 2002), averaged over 300 percent. Thus, even taking methodological differences into account, 50 percent is probably a lower bound hurdle rate based on the internal rate of return measure.

[43] Source: Graham and Harvey [2001]

The average accounting return method does not use cash flow (the true measure of the project's impact) as do the NPV and IRR methods. Instead, it uses an accounting version of net income and relates this measure to the book value of the investment over the life of the study period. An even more serious drawback is its failure to account for the pattern (timing) of net benefits. Finally, the profitability index is the ratio of the present value of future expected cash flows *after* initial investment divided by the amount of the initial investment. This metric does not provide a complete picture of the return on an investment, but it has some utility when a budget constraint is being imposed by providing an indicator of the efficiency of cash flow generation per dollar of investment.[44]

In summary, the NPV method provides a direct approach to ranking the merits of R&D projects, providing the projects being compared are approximately the same size. Much of corporate finance is concerned with comparing mutually exclusive projects. In such a situation, the NPV method still is direct and straightforward. The IRR method, on the other hand, is independent of the MARR, which is an attractive feature, so rankings will not be reversed by significant changes in the discount rate. However, this method requires complicated incremental analysis to achieve similar rankings with NPV.[45] When comparing independent projects (essentially the case for retrospective studies), neither method seems preferable because the pattern of net benefits (affected by the length of the study period) determines relative rankings.

2.7. Estimation of Costs and Benefits

Quantitative estimation of costs and benefits using metrics and impact measures discussed in previous sections are often the only information produced in an economic impact study. This is unfortunate because much information useful for strategic planning and mission modification can be gained from qualitative analyses. Furthermore, the two types of information can be more efficiently obtained simultaneously during data collection activities.

Nevertheless, quantitative impact assessments have unique and significant effects on decision making and provide a level of validity to the overall analysis that cannot be otherwise obtained. However, properly constructing and executing quantitative impact assessments is difficult and is frequently done incorrectly. This section provides some guidelines for matching quantitative analysis approaches with the nature of the research program or project being studied and the projected availability of impact data of varying quality.

2.7.1. Determination of Costs and Benefits. The objective of an economic impact study of a government R&D program is to estimate aggregate (social) economic impacts. Therefore, the aggregate benefits from the project must be compared with the aggregate costs independent of what is replaced by that project through user fees (i.e., independent of substitution of costs between government and industry). The resulting rate-of-return

[44] Ross, Westerfield, and Jaffe [2002, Chap. 6].

[45] Au and Au [1992, pp. 30–31].

calculation is compared with a "hurdle rate" to determine if *ex ante* the project should be undertaken and *ex post* if the project was worthwhile.

Obviously, the designation of costs and benefits and the subsequent use of these data in constructing outcome metrics are key steps in the quantitative portion of the economic impact analysis. Unfortunately, the methods for selecting, calculating and interpreting the appropriate metrics are not rigorously developed. In this section, problems that frequently arise with respect to definition and identification of appropriate benefits and costs are discussed.

As an initial step, the analyst must identify the objectives of the research program and then carefully describe the technical outputs (Section 2.2.4). These outputs constitute the technological infrastructure that government contributes to an industry's overall technology. Thus, they are the drivers behind the eventual economic impacts. This analysis is facilitated by placing the roles of the technology infrastructure in the context of the relevant industries' technology base and competitive strategies. Understanding how this infrastructure complements the industry-developed proprietary technology elements is essential to the eventual selection of metrics, impact hypotheses, and construction of the survey instrument. Moreover, relating industry demand for the infrastructure to private sector investment behavior will capture the market failure rationale for the government research program and thereby facilitate the eventual presentation of the study's results to stakeholders.[46]

All benefits and costs need to be identified and correctly labeled. Mislabeling of benefits and costs as social or private can occur when exchanges of funds take place between the government R&D agency and industry in the course of developing and transferring technical infrastructure. As discussed in Section 2.4, such mislabeling is more of a problem for the calculation of benefit-cost ratios, although such errors will obviously affect other metrics as well.

In that example—a NIST study of the economic impacts of standard reference materials (SRMs) for sulfur in fossil fuels (a major regulatory concern)—a debate ensued over how to treat industry purchases of the SRMs (Martin et al [2000]). One option was to view them as a negative cost and subtracted these costs from the other costs (development costs) incurred by NIST. However, because an SRR was being estimated, a sector transfer of costs is not relevant. That is, the total social costs (developing and producing the SRMs) should remain the same, regardless of who pays them. By subtracting this "negative benefit" from NIST (social) costs instead of netting against gross benefits, a higher rate of return and benefit-cost ratio were obtained. In general, industry operating costs, including assimilation ("pull") costs, are subtracted from gross benefits realized by industry to get a net benefits measure for the numerator of the impact ratio.

[46] A persistent problem in influencing science and technology (S&T) policies is the difficulty in explaining why and how the government role differs from the private sector role. Hence, presenting a qualitative analysis of the government's contribution of certain types of technological infrastructure in the context of industry competitive dynamics helps reveal the significant differences in private-sector investment incentives among the major technology elements identified in Fig. 3.

A parallel and closely related step in the analysis process is the development of impact hypotheses. An impact hypothesis formalizes the proposed relationship between a set of inputs, including the infrastructure elements of the industry's technology, and outcomes. If fully implemented, taking this extra step beyond simply identifying several benefit variables can aid subsequent qualitative analysis by relating the government's research program's impacts to the broader structure of economic activity.

Impact hypotheses should be formulated and clearly stated to serve as guides or forcing functions for the data collection. The statement of impact hypotheses helps to prioritize metrics and to focus the analyst on the relationships among metrics. The latter is important for designing data collection instruments and in the actual conduct of data collection, especially when direct interactions with respondents are part of the survey approach.

However, the analyst may have to tradeoff the feasibility and cost effectiveness of alternative metrics derived from the hypotheses against needs of target audiences. As discussed in Section 3, the effectiveness of data collection varies significantly across industries for a number of reasons and different metrics have different costs associated with data collection independent of the response characteristics of the target population.

2.7.2. Extrapolation and Characterization of Net Benefits. As described in Section 2.2.2, government research programs can exist for a number of years, followed by additional time for marketplace diffusion and ultimate impact. The timing of an economic impact study therefore presents a difficult problem. The significant time required for diffusion and impact of major elements of an industrial technology, including technology infrastructure, must be traded off against the perishable nature of impact data. These factors coupled with occasional political considerations often influence the analyst to undertake impact assessments at intermediate points in the technology life cycle. To obtain complete (full life cycle) impact estimates, such decisions require extrapolation of benefit estimates in both time and industry space.

For any time period selected for economic impact analysis, outcome/benefit data often are difficult to explicitly estimate on an annual basis. Usually, industry respondents will provide an estimate of average benefits received over a specific time interval or specific benefit estimates for just a few years of the projected impact time series. If discrete values for annual net benefits cannot be estimated, the analyst must still specify an impact period but cannot use measures such as rates of return, which require well-defined time series. This lack of knowledge about the life cycle pattern of net benefits constrains extrapolation, as well.

Even when industry respondents forecast continued benefits into the future and the time series is reasonably well specified up to the point of the study, the analyst still has a decision to make regarding extrapolation. Usually, the analyst will truncate industry estimates of future benefits to reduce extrapolation to a reasonable range (one that is compatible with a conservative estimate of the remaining portion of the technology life cycle).

Cross section (industry space) extrapolations are also a methodological issue and frequently must be addressed for the industry being surveyed. That is, only a portion of the companies respond to surveys with usable information for the set of desired impact

metrics. In such cases, the analyst can extrapolate in the industry space using benefit data obtained from the responding companies and weighted (if appropriate) for the size of the non-responding firms. However, extrapolation in industry space is not always justified. If the benefit estimates vary significantly among responding firms of the same size or simply differ significantly for no apparent reason, confidence in extrapolation may not be sufficient to allow this approach.

For example, the study of NIST's Cholesterol Measurement Program surveyed 17 cholesterol measurement instrument manufacturers and diagnostic chemical vendors (Leech [2000]). Seven responded completely and five provided some usable quantitative data. The analyst decided that the two additional companies that identified value from NIST but did not provide quantitative benefit estimates should not have benefits imputed to them because of the heterogeneity of the sample.

Finally, in cases when extrapolation is not made or when quantitative estimates can be obtained for only a portion of the types of benefits actually realized by the respondents, a common approach among analysts is to characterize such impact estimates as "conservative" or "lower-bound". The remaining categories are described in qualitative terms. However, such characterizations presume that the estimates and descriptions themselves are accurate for the portion of the impacted population studied and that significant net benefits would have been found if the remaining companies had been studied. In fact, estimates can overvalue the impact on the particular segment of the population for which impact data are obtained. Where an overestimate occurs, the resulting measure of impact is definitely not conservative for the population (industry or industries) studied and could overwhelm any additional net benefit attributable to other segments. Such a possibility is one of several reasons for undertaking sensitivity analyses.

3.0 Data Collection Strategies

3.1. General Strategies

The data needed for an impact study do not typically reside in industry or within a government agency's accounting structure. Therefore, the analyst must work with both industry and the R&D agency to define, collect, and refine the required benefit and cost data. Issues such as imputation, extrapolation, and averaging continually arise when working with survey data. For the most part, companies report information that is available in accounting and other business information tracking systems. NIST's considerable experience with economic impact assessments indicates that companies almost never generate new information solely for reporting purposes. Thus, even when reporting is required by regulation, the quality of the data must be carefully assessed as a basis for selecting the appropriate metric set.

Because data collection is typically labor intensive and hence both time consuming and expensive, careful consideration must be given to the set of industries targeted for surveys. As previously discussed, the nature of the economic study (prospective or retrospective) and the projected industries in a supply chain that are directly impacted are critical factors in determining the study's scope.

3.2. Data Quality Issues

Industries react differently to solicitations for participation in the conduct of an economic impact study. Because response rates and levels of effort on the part of respondents significantly affect the quantity and quality of the impact data, extreme care must be taken when developing collection strategies. For example, the needed data frequently must be "constructed" by obtaining a consensus and/or composite estimate from several people within the same company. This can be a difficult and frustrating process, frequently resulting in skewed quality of data across companies in one industry and even more so across industries in the supply chain under study.[47] Whatever, the selected approach, its strengths and weaknesses must be taken into account during the analysis stage of the study and documented in the final report.

Published data occasionally can be used in place of, or in addition to, industry surveys. In a study (Leech and Chinworth [2001]) of support provided by NIST for the key electronic funds transfer standard (the Data Encryption Standard), conventional surveys were taken of the relevant hardware and software vendors. However, the use of published Federal Reserve Board data for a case study of impacts on financial services turned out to be a better approach than a survey of banking institutions. The Fed data were reasonably good for purposes of the desired analysis, while the banks' accounting systems were not structured to directly yield the desired impact measures and the long time series used in the study made the identification of and access to reliable data sources difficult.

The same general set of issues exists with respect to the cost data required from the R&D agency. Government R&D program cost records are typically poorly matched with the boundaries of the study because agency accounting systems are not designed with impact assessment in mind. Specifically, most cost centers are not structured along programmatic lines, at least not in ways that match the project elements delivered to specific industries to solve specific problems (which define the boundaries of economic impact studies). Moreover, accounting procedures change over time, imparting discontinuities to the cost time series.

Therefore, because of the need to match benefit metrics with available cost data, assessment of cost data quality should start as early as possible within the time frame of the study. In fact, one lesson from the NIST economic impact studies is that a thorough assessment of the quality of available cost data should be made *before* expending the considerable resources required to obtain outcome (benefit) data from industry. That is, the quality of the cost data influences the feasible types of benefit data to be sought from industry.

Matching analytical techniques with the amount and type of data collected is also an important quality issue. Contractors must spell out data manipulation and statistical computation techniques, so those aspects of the analysis can be checked. More generally, the final report must clearly state and respond to the analyst's assessment of the quality of the data. This objective can be achieved through the use of interval estimates, sensitivity

[47] See Trajtenberg [1990, Chap. 1 and pp. 164-169] for an example of the difficulties of obtaining data across the complete set of hypothesized beneficiaries.

analysis, and general discussions of the data set. For example, the implications of dependency of the estimates on the input from a single large firm that dominates the market being studied or the degree of success in accessing the desired multiple sources within responding companies should be explicitly recognized and assessed.

3.3. Data Collection Instruments

A number of techniques for data collection can be used:

Site Visits. Usually provide the greatest breadth and depth of information, as they allow for extended questioning of several individuals within the company with complementary information. This breadth of interaction, including direct observation of the responding company's overall business and approach to the topic under study are advantages. In fact, this technique is often an essential step in defining metrics and impact hypotheses, as well as pre-testing draft survey instruments. However, the information supplied will be largely qualitative unless interview guides with specific quantitative questions are submitted ahead of the interview date. Even in such cases, this approach is not well-suited to quantitative data acquisition because it is time consuming and very labor intensive, making it the most expensive form of data collection. Focus groups are a variant.

Telephone Interviews. A low-cost substitute for on-site interviews. Scheduling the call and providing the respondent with an interview guide ahead of time are essential. This technique allows more respondents to be reached but requires a more structured and refined survey instrument.

Written Survey Instruments. Relatively low-cost and broad-coverage approach. Allows maximum consistency of responses. Because the information requested is not collected in real time, this is the best mode for acquiring data that have to be looked up or verified and approved by company officials for release (typical of quantitative data requests). Formats for written surveys can be traditional mailed paper copies of the instrument, sent as email attachments, or Internet-based.

Several other factors besides the format influence quantity and quality of data collection activity:

- The pre-test is an essential step for any type of structured survey. A critical part of this step is to refine hypotheses/metrics and hence specific questions.
- For most survey approaches, response rates are highly dependent on pre-survey contacts, including requests to industry for participation from the sponsoring R&D agency.
- A key issue is the development of sampling strategies to not only ensure the acquisition of data in the needed categories of information, but also to enable extrapolation to the industry, sector, or national levels.

Written surveys are by far the most cost effective approach to data collection for sampling large populations. However, these approaches will yield poor results or will not work at all in some cases, especially if they are received without prior introduction or sent to a "list" that is not targeted to the appropriate respondents/gatekeepers within target organizations.

Another serious weakness of printed survey instruments is the inability to iterate questions and answers with participants. Industry respondents are often unfamiliar with the economic issues behind the survey. Respondents identified by the R&D agency typically have the same technical orientation as the government project staff and therefore frequently have difficulty understanding economic impact questions. Therefore, complementary telephone contact is often required. The interviewer can then explain the question until the “language” is right for eliciting a complete and accurate response. The interviewer can also determine if a second or even a third individual within the responding company needs to be contacted for participation in the response. As a result, the data collection plan for most government R&D impact studies should include telephone contact information. Constructing and forwarding at least an interview guide, if not the actual survey instrument, prior to the scheduled telephone interview facilitates such interactions.

One limitation of the personal contact approach is the reluctance on the part of a significant number of companies to have information conveyed verbally to outsiders. Also, technical staff often cannot answer at least some of the economic/financial questions and must refer the interviewer to appropriate individuals elsewhere in the company, thereby further complicating company control and extending this more labor-intensive approach to data collection.

For these reasons, many companies prefer, even insist, on a written response. This typically means identifying a “gatekeeper” within the company who solicits inputs from appropriate individuals and clears the response. Success with this latter approach requires some initial telephone communication to ensure an appropriate gatekeeper has been identified and that the substance of the questions is understood. Criteria for what constitutes an “appropriate gatekeeper” include knowledge of the R&D agency, contacts within the company necessary to fully complete the survey, authority or access to authority to permit target respondents to participate, and a commitment to the usefulness and adequacy of the survey instrument.

3.4. Factors Affecting Successful Data Collection

Because less effort is required, many analysts make only a single contact with potential respondents within the target population. This usually entails mailing a survey unannounced to potential respondents and hoping for a decent response rate. Such an all-or-nothing approach frequently yields low response rates, in part due to targeting of the wrong individuals within surveyed companies. The latter effect lowers the quality of what data are obtained. As a result, the population either is re-surveyed at additional cost or a low-quality database is accepted.

As indicated in the previous section, successful data collection requires a multi-step approach, with preliminary (test) surveys early in the project guiding hypothesis development and industry coverage, as well as construction of the final survey instrument. Obtaining economic impact data is greatly facilitated by established relationships between the R&D agency and industry. Agency staff can often provide essential contact lists and thereby facilitate identification of accurate survey populations. These relationships also tend to promote higher levels of cooperation by the industry responding to the survey. In the lean corporate organization of today, industry personnel

are more pressed for time than ever before. Consequently, getting their attention can be quite difficult. An impact study team member is just another outsider from industry's perspective, without a request from the R&D agency to cooperate with the survey. The difference in response rates with and without the agency's assistance can be significant.

Unfortunately, the R&D agency's contacts are often limited to one industry. Access to only one of the several industries with hypothesized significant direct benefits constrains the desired data collection effort for the relevant supply chain. The agency's contacts are also frequently limited to one person within a company. A technical person within a target company may be willing to respond to the survey, but often does not have all the required information (especially the critical financial information) and must therefore approach other employees who are not familiar with the R&D agency and tend to be less cooperative. Moreover, the responding individual frequently must comply with a company policy requiring approval from higher levels of management. These managers may have little or no appreciation for the agency's contribution and therefore choose the low-risk and time-conserving option of not cooperating.

The only way to mitigate such constraints is for the R&D agency to cultivate relationships at higher management levels in the companies targeted for data collection. This approach typically cannot be done in a short period of time. In fact, such relationships need to be built during the strategic planning phase of the R&D project, when commitments to objectives and of resources are made.

Another factor in successful data collection is confidentiality. Industry has a strong sensitivity to any information given by employees to outside entities. Therefore, attaining high response rates requires imparting confidence to the individuals and companies being interviewed that proprietary information will not be disclosed. With respect to qualitative information, industry respondents assume that anything they say in terms of opinions, assessments, etc. may show up in print, so the analyst must provide assurances that appropriate discretion will be used with respect to attribution (for example, only quoting anonymous sources). Moreover, industry may assume that only the contractor will handle and see individual responses, so this condition should be made clear to respondents.[48]

Rather than rely on assumptions by respondents, an R&D agency's policy with respect to contacts and exchanges of information should make sure that industry realizes what is meant by a promise of confidentiality for the purposes of data collection and analysis.

[48] Two laws are relevant. The Trade Secrets Act [18 U.S. C., Section 1905] bars Federal government employees from releasing trade secrets, processes, operations, style of work or apparatus as well as the identity, confidential statistical data, amount or source of any income, profits, losses, or expenditures of any person, firm, partnership, corporation or association. The Freedom of Information Act [5 U.S. C., Section 552(b)] exempts trade secrets and commercial or financial information from disclosure via a FOIA request.

4.0 Retrospective Economic Impact Studies

4.1. Selection of Research Projects for Economic Analysis

As part of a program evaluation exercise, individual research projects are studied to generate performance data. However, the number of projects that can be assessed at a point in time is limited by the cost of such studies, the time required to complete each analysis, and the maturity of each project when case study selection is made.[49]

Sometimes research projects can be studied independently of a broader program evaluation. NIST began economic impact studies of its laboratory (infratechnology) research programs in the early 1990s for several reasons: (1) its mission was not being questioned but neither was it well understood by policy makers, and (2) the individual research programs were not particularly visible to both public sector and industry stakeholders. As a result, NIST's long-term budget history was not good.

In response, retrospective impact analyses were initiated not only to inform management on relative performance across projects, but also to develop a database of quantitative and qualitative impact information that (1) educated the policy and budget processes on the types and magnitudes of economic impacts and (2) imparted credibility to the agency's overall mission. This effort has been successful with respect to the second objective, but achieving the first one requires more than retrospective impact data as previously discussed.[50]

For these last two objectives, the subjects of economic impact studies are not chosen randomly. Rather, topics are selected based on the R&D agency's perceptions that significant economic impact had occurred, thereby affording the opportunity to document both the types and magnitudes of the economic contributions being realized from laboratory research programs and projects. However, managers' perceptions are not always accurate with respect to the types of impacts and the absolute and relative magnitudes of these impacts. Nevertheless, in the majority of cases, they seem to be able to identify the higher impact areas of the R&D agency's research and associated services.

By selecting studies that cover a range of technologies and their associated industries, the R&D agency's management can acquire a rich database on how the

- agency's research affects the three major stages of economic activity (R&D, production, and commercialization)[51],
- mechanisms by which the technical knowledge produced by the agency is transferred to industry and the resulting impacts occur, and
- magnitudes of these impacts are realized in total and over time.

---

[49] In any each year only a fraction of a program's projects will be at a stage that meet the criteria discussed earlier for initiating an impact assessment.

[50] See, for example, Figure 2 and the associated discussion in Section 1 and Section 5.

[51] The scope of the term "commercialization" is much broader than the initial market introduction (innovation). It includes the process of market penetration, including such activities as after-sales service.

NIST's ATP had a somewhat different motivation for its impact assessment program. ATP's mission has been continually attacked since the Program's creation in 1988. Thus, impact data are essential to show that the mission is achieving significant economic impacts and thereby provide positive feedback into the mission rationale debate (see Fig. 2). Because a need here was to rationalize the overall mission, data summarizing program-level performance are essential.[52]

In general, only a fairly expansive level of effort, including studies of projects that cover the major technological areas targeted by the R&D agency, can provide an effective economic perspective on the nature and impacts of an agency's research programs, including information relevant for role justification/modification and strategic planning.

Selection of projects for impact assessment is implemented with a careful screening process in order to avoid disappointment with respect to the eventual attainment of useful impact data. Topics for impact assessment are selected jointly by the analysis team and participating operating units. The screening process is initiated by having the managers of these units submit short lists of candidate projects, which are assessed using the following criteria:

(1) Has the project been transferring technical knowledge to industry for at least several years?

(2) Does the project appear to have had substantial economic impact?

(3) Does the agency have direct contact with the industries believed to be the primary beneficiaries and is there reasonable expectation that contacts in these industries will agree to respond to a survey?

(4) Does the agency have cost data available that closely match the research and technology transfer activities conducted under these projects?

4.2. Estimation and Interpretation of Quantitative Impact Estimates

Over 30 NIST infratechnology research programs or projects have been analyzed for economic impacts over the period 1992-2002. Of these, over 20 are deemed to have achieved sufficient coverage of industry impacts to allow some comparisons with one another and to a lesser degree with impact studies of industry R&D projects done by economists and business analysts over the past several decades. Ideally, comparisons with industry impact studies would allow a rough assessment of the relative efficiency of public vs. private R&D.

As described in Section 2, these studies rely primarily on counterfactual hypotheses, even though some impact data are frequently available for the period before the government intervention. Without the technical infrastructure provided, industry typically incurs additional costs due to uncertainties about measurement, testing, calibrations, technical data development and qualification, interface design, quality assurance procedures, etc. To compensate for this uncertainty, companies either perform these activities more frequently, procure services from multiple vendors to serve segmented

[52] See Tassey [forthcoming] for the types of data necessary to address the mission rationale issue.

markets, or over design products to ensure performance specifications are met. In all cases, costs to the affected industries are increased. Cost reductions from the government research program typically constitute the net economic benefits.

Table 6

Retrospective Economic Impact Assessments: Outputs and Outcomes of NIST Laboratory Research, 2000–2002

Technology/Project	Output	Outcomes	Measure
Photonics: power and energy calibration	• Calibrations	• Increase productivity • Reduce transaction costs	IRR: 43–136% BCR: 3–11 NPV: $48M
Chemicals: standards for sulfur in fossil fuels	• Measurement methods • Reference materials	• Increase R&D Efficiency • Increase productivity • Reduce transaction costs	IRR: 1,056% BCR: 113 NPV: $409M
Semiconductors: Josephson volt standard	• Measurement methods • Reference materials	• Increase R&D efficiency • Enable new markets	IRR: 877% BCR: 5 NPV: $42M
Quality Techniques: National Quality Award	• Quality improvement techniques	• Improve product/service attributes • Increase productivity	IRR: N.A. BCR: 207 NPV: $25B[53]
Communications: data encryption standard	• Standard (DES) • Conformance test methods	• Accelerate new markets • Increase R&D efficiency	IRR: 270% BCR: 58–145 NPV: $345M–$1.2B
Communications: role-based access control	• Generic technology • Reference models	• Enable new markets • Increase R&D efficiency	IRR: 29–44% BCR: 43–99 NPV: $59–138M
Energy: gas standards for regulatory compliance	• Standard (NTRM)	• Increase productivity • Reduce transaction costs	IRR: 221–228% BCR: 21–27 NPV: $49–63M
Manufacturing: product design data standard	• Standard (STEP) • Conformance test methods/facilities	• Increase R&D efficiency • Reduce transaction costs	IRR: 32% BCR: 8 NPV: $180M

SRR=Social Rate of Return, BCR=Benefit-Cost Ratio and NPV=Net Present Value

Table 6 provides a summary of the results from nine recent impact studies. Most of these studies and earlier studies exceed the approximate internal hurdle rate of a 50 percent IRR, which is considered to be an upper bound based on economic studies of social rates of returns.[54] Many of them were significantly above this hurdle rate. The few

[53] Extrapolated to the national (economy-wide) level. This distinguishes the study from the other retrospective assessments undertaken of NIST's infratechnology research, which only estimated benefits for the initial market or markets affected by the NIST program (i.e., only markets for which primary data sources could be accessed).

[54] Selection of this approximate hurdle rate is based on "social" rates of return for industrial innovations estimated in NSF-sponsored studies by Mansfield [1977], Tewksbury et al [1980] and in other studies summarized by Griliches [1988, 1995] and Mansfield [1991]. Griliches judged the social rates of returns for industry R&D to be in an interval of 20–50 percent and Mansfield states that the rate of return from industrial technology R&D "has been very high, frequently 40 percent or more". Thus, choosing 50 percent as a hurdle rate would seem conservative. However, one of the most intensive and thorough

that did not exceed this hurdle rate still had SRRs in the 25 to 35 percent range, which exceeds the lower end of the hurdle rate interval based on industrial R&D studies. Once again, the variances across studies in terms of scope, quality of data, and methodology reduce the ability to compare results.

The large number of studies across many different industries with SRRs significantly above any reasonable hurdle rate implies that, even with allowances for differences in the major factors determining impact estimates, the returns to government investments in infratechnologies have been above average. However, these studies were not randomly selected for reasons previously stated. In fact, an analysis of a random selection of NIST infratechnology research and services programs would likely yield a lower average rate-of-return estimate. Nevertheless, the relatively large number of economic studies completed covering a wide range of technologies and industries provides strong support for the role of infratechnologies in contributing to the overall efficiency of technology development and use in a modern economy. Certainly compared to other R&D agencies, these studies represent a large and diverse database of economic impact assessments and thereby provide relatively strong support for NIST's infratechnology research mission.

For reasons described in the next section, significant differences exist even among government projects of the same general type (infratechnologies and generic technologies) and between these projects and the typical industry project, so that only limited comparisons are possible. Moreover, academic studies have varied in methodology and quality of data collection, further weakening legitimate comparisons.[55]

In summary, the cumulative effect of these retrospective studies on the policy and budget processes seems to be one of validating the NIST infratechnology mission. When combined with prospective studies (discussed in Section 5), economic analyses promote more favorable responses to program/project initiatives from the policy and budget processes.

4.3. Comparing and Interpreting Economic Impact Data

Microeconomic impact studies of government research programs/projects can be compared to a degree, if certain requirements are met. These requirements include

- equality of coverage by the impact analysis with respect to the projects making up a research program
- equality of coverage within industries and across supply chains
- similar definitions of metrics (costs and benefits)
- similar quality of impact data
- a comparable point in technology life cycle when analysis is undertaken

---

studies of R&D investment for a single industrial technology (CT scanners), yielded a much higher social rate of return estimate of 270 percent (Trajtenberg [1990, p. 167]).

[55] See Griliches [1995] for a summary and assessment of these latter groups of studies.

An analysis of the NIST retrospective studies indicates that a significant portion of the variation in estimated economic impacts has been due to differences in the portion of hypothesized impacts for which impact data could be obtained from the relevant industry supply chain. Specifically, coverage of the relevant supply chain varies, as does coverage of firms within a particular level/tier (industry) in that supply chain. In general, extrapolation potential is limited, so coverage differentials remain significant.

Further, assessments of individual projects should be viewed in the broader context of the set of related projects (within the same research program). This is because economies of scope often exist for R&D projects addressing several categories of related technological infrastructure. Thus, one project with a modest impact estimate may, in fact, enable other types of projects to achieve higher impacts. For example, many NIST laboratory program areas develop measurement methods and then supply standard reference materials (SRMs) and calibration services (through separate projects) to help industry assimilate the measurement infratechnology. The same scientific and engineering skills along with research facilities can be applied to several project categories. Therefore, studying just one project within a research program can sometimes produce misleading results and compromise comparisons across studies.

Similarly, NIST's ATP has to varying degrees over its existence emphasized focused funding of technological areas with perceived significant economic potential but large market failure problems. This strategy means funding multiple projects in the same technological area, as opposed a distributed or general funding strategy (each project viewed independently). Impacts assessments can analyze projects in isolation but should also consider their potential role as to contribute to a broader programmatic objective (the focus area).

Another reason that argues for discouraging comparisons is the significant differences among programs or projects studied with respect to the ability of the analyst to obtain benefit data from industries beyond the tier or tiers in the supply chain that initially or directly receive economic benefits from the NIST program. As mentioned above, the conventional definition of the social rate of return includes estimates of economic benefits from both sides (sellers and buyers) of the market for a technology. In some of the NIST studies, only the supply (seller) side could be surveyed effectively. In a few others, data were obtained not only from sellers and buyers in the initial market to which the NIST technology infrastructure was delivered but also from an adjacent industry in the relevant supply chain. Such differences obviously affect measured economic impacts.[56]

Similarly, the ability to obtain accurate costs varies across studies. As pointed out, government cost accounting systems do not closely match programmatic content. Moreover, research and transfer costs are frequently shared. In such cases, cost trends can be analyzed in terms of the relative contributions of government and industry, but accounting for the respective contributions can be difficult. For example, several of the

[56] However, as discussed in Section 2, effective management of the government research project includes sound strategic planning, a part of which is the establishment of working relationships with all target industries (defined as those industries that can benefits significantly from the research project's output). Failure to do so will lead to lower economic impacts.

NIST projects analyzed were co-funded by industry, but the funding came in different forms (internal research, direct funding for government research, funding for research consortia, indirect funding through certification and reference materials fees, contributions to technology diffusion or standardization, etc.). Coupled with the fact that multiple projects are often funded under the same research program but by different parties for each project further complicates cost estimation.

The inadequacies of the metrics themselves provide another reason for limiting comparisons. As discussed in Section 2.3, the commonly used impact measures each have attractive and unattractive features. As a result, NIST impact studies now compile estimates of the three most commonly used measures: internal rate of return, benefit-cost ratio, and net present value. Attempts to compare research programs using these metrics can lead to confusion because variations in net benefit time series constructed in typical impact studies can result in inconsistent rankings of programs or projects by the three measures. The best approach to interpretation is to use all three metrics together with an understanding of the strengths and weaknesses of each.

As an example of this point, consider the fact that the majority of NIST *projects* are small in terms of impact. Hence, while many are easily justified on a rate-of-return or benefit-cost basis, the absolute economic value (net present value) is typically small. The photonics calibrations services program (Marx *et al* [2000]) is typical in this regard. Three individual projects were studied as part of the economic impact assessment. For each project, the results show an excellent internal rate of return and a good benefit-cost ratio (hence each project is justified). However, individual projects delivered only modest NPVs (for example, the 248 nm project yielded an NPV of $4.7 million).

NPVs at the *program* level are, of course, larger. In this case, the three projects studied in the photonics research program studied had a combined NPV of $48 million (first row of Table 6). Occasionally, a technological infrastructure program such as the set of standards for sulfur in fossil fuels (second row in Table 6) supports several large industries in a supply chain and the economic magnitude of this support is measurable over a relatively long period of time. In such cases, the economic impact is quite large (estimated NPV was $409 million for this program).

As stated above, research programs benefit from economies of scope across projects. The management implication is that projects should be managed as part of a broader program with content and timing geared to the broader research program's objectives. From the analyst's perspective, the important point is that sometimes the best approach is to analyze several closely related projects as a group (sulfur SRMs), while in other cases (photonics calibration services) it is better to analyze related projects separately. Thus, care is required when comparing NPVs across studies because of differences in programmatic scope.

Efficiency measures (benefit cost ratios and rates of return) can be compared across programs of different sizes because they are ratios as opposed to absolute values. However, even with these latter measures, caution is in order. For example, an examination of the sample of impact studies in Table 6 indicates only a modest correlation between the BCR and SRR measures. This lack of correlation results from the

differences in the way the two measures are calculated (Section 2.3). In particular, the pattern of net benefits over the time period studied affects them differently.

An important point with respect to interpretation is to not rely just on the quantitative estimates to characterize economic impact. Rather, qualitative and quantitative information should be integrated into a broad analysis of the project's impact. Instead of simply presenting bottom-line impact estimates, the patterns of benefits of the study period can be analyzed in the context of the interactions of the technology trends and the competitive dynamics of the relevant industries. For example, the economic impact of product innovations will usually be less if significant process innovations are not made. Conversely, the demand for process innovations is totally dependent on the uniqueness of product innovation. The impacts of infratechnologies and related standards, as well as the timing of these impacts, depend on the patterns of product and process innovation.

In summary, providing such qualitative complements to the quantitative impact estimates provides a real-world picture of the government project's impact and also reflects on the rationale for government doing the project.

4.4. Factors Affecting Measurement of Economic Impact

Technical infrastructure, including generic technologies and infratechnologies, is intricately integrated with the core competence and market strategies of an industry and, more broadly, of entire supply chains. Thus, corporate strategies across several industries and the resulting marketplace dynamics determine the need for government R&D support and eventually the patterns of economic impact. Prospective (strategic planning studies) obviously provides valuable information on the number of target industries and optimal targets for research projects in a particular area of technology and therefore provide the basis for resource allocation and program management.

However, retrospective studies also can provide valuable insights into these relationships and in so doing offer valuable guidance for both effective management of R&D programs and the conduct of strategic planning. Unfortunately, the ability of economic impact analysis to access and to obtain useful impact data varies significantly across studies, and the amount and quality of impact data obtained directly affects a study's results. Managers of government R&D programs can both increase the economic impacts of their research projects and enhance the ability to measure those impacts by understanding the relationships between corporate strategy, industry structure, technical infrastructure, and the institutional levers within companies that affect these relationships.

4.4.1. Underestimation due to Incomplete Coverage of Impacted Industries. A majority of government investments in technology infrastructure impact several industries because economies of scope are present. For example, interface standards allow component suppliers to compete in the downstream industry whose firms provide the system technology requiring the component. Component suppliers benefit because requirements for accessing the market (interface specifications) are well defined, thereby reducing uncertainty and projected product cost with respect to a market entry decision. The system integrators benefit because they can choose efficiently among several suppliers. Similarly, acceptance testing standards allow two sides of a market transaction (i.e., two different industries) to conduct business efficiently (low transaction costs).

Scope determinations should therefore always be a central part of program planning and project selection because the broader the scope of potential utilization of the resulting infrastructure, the greater the economic benefits realized from the infratechnology research project.

Moreover, trends in industry structure require a multi-industry strategic planning focus by government laboratories that provide technical infrastructure. Such a “supply chain” focus for research program planning is increasing for a number of reasons. For example, virtual as opposed to actual integration of industries in the typical high-tech supply chain has increased the need for open systems with associated interface standards. The trend toward vertical disintegration creates more market place transactions and therefore places increased demands on technical infrastructure for demonstrating compliance with performance specifications. Also, greater distribution of R&D across these supply chains requires multi-industry infrastructure support and underscores the need for involving several industries in strategic planning exercises. The opportunities for synergies (economies of scope) have therefore increased for many NIST research programs and therefore so have the demands for a broader scope to strategic planning.

An example is NIST’s work in calibration services for laser power meters. The planning focus involved interactions with power meter and laser manufacturers, but not the downstream industries that use these devices (lithography equipment and semiconductor manufacturers). Without infrastructure support, semiconductor manufacturers over time evolved a relatively crude empirical approach to calibration that would seem to warrant improvement.[57] One can speculate that interactions with these latter two industries would have increased diffusion of the NIST work and raised the economic impact from an already good result.

In contrast, a NIST project involving mathematical modeling for software used in assessing performance of a particular class of semiconductor devices (IGBTs) is an example of complete supply chain coverage. Three levels (industries) in the supply chain for automotive ignition systems were directly targeted: software vendors, semiconductor component manufacturers, and equipment manufacturers (Gallaher et al [2002]). The objective of the NIST project was to develop software that simulates the performance of this class of semiconductors in a system environment. The resulting design automation increased the efficiency of industry R&D by both developers of the semiconductor chips and manufacturers of the products using the chips. Companies at all three levels participated in testing the simulation modeling software. Relatively few firms were involved at each of the three tiers, but any involvement greatly increases the subsequent diffusion of a project’s technical output to other firms in each tier of the supply chain.

However, analysis of these and other NIST research projects indicates that even when multiple industries are directly involved with NIST, particular measurement infratechnologies within industries seem to get targeted from among a broad set of opportunities without the benefit of prior planning and hence prioritization. The IGBT project targeted one of a number of semiconductor device classes. Similarly, laser “dose”

[57] One piece of equipment on a production line is selected and the laser source is fiddled with until the image quality is acceptable. A power detector is then used to measure the source, and the other pieces of equipment on the production line are calibrated using that detector.

control (calibration of laser power meters) is one of about 10 key process variables critical to semiconductor manufacturing, most of which do not have NIST-supported standards. No systematic planning process was apparent by which these projects were selected from the relevant opportunity sets. Thus, even though the two projects achieved admirable economic impacts, even greater impact opportunities were likely missed by not systematically assessing project opportunities.

In at least one study, NIST's electrical power and energy calibration services (Link [1995]), the largest single economic impact came from a totally unanticipated level in the supply chain: reduction in litigation costs incurred by electric utilities and state regulatory commissions resulting from traceability to the national watt hour standard maintained by NIST. Industry ranked this impact the highest when the hypothesized greatest impact was a reduction in measurement uncertainty. The magnitude of this previously unobserved regulatory need uncovered by the analyst conducting the study caused NIST's management to increase emphasis on getting calibrations to the relevant industries faster. Identifying such potential impacts *a priori* is difficult, but certainly cannot be achieved without a strategic planning exercise covering the entire supply chain and, in this case, the associated regulatory environment.

Several of the NIST studies provide examples representative of areas of research that affect industries *across* supply chains. One of the reasons that the sulfur-in-fossil-fuels SRM program yielded such high economic benefits is that the same measurement technique was applied to a large number of related SRMs, which significantly affected several large industries over time. Moreover, the underlying measurement method (IDMS) was sufficiently sophisticated that industry respondents voiced the opinion it likely would not have been developed at a later date by another source. This assessment led to an unusually long estimated period of benefits (17 years). More often, a counterfactual hypothesis is accepted based on industry interviews that some combination of industry/university sources would have replicated the NIST infratechnology at some point in the future, resulting in a truncation of the net benefit time series.

4.4.2. Underestimation Due to Inadequate Data Collection. Underestimation can result from inherent difficulties in obtaining impact data due to the nature of certain industry structures. As a result, several levels in a supply chain can be targeted and economic impacts presumably achieved without significant portions of these impacts being measured in an impact study. This situation occurs because the ability to identify and measure economic impacts varies significantly, depending on the structure of the relevant industries and the government R&D agency's relationships with these industries.

For example, a study of NIST's thermocouple calibration program (provides accuracy in temperature measurement) was largely relegated to one industry—a group of small firms that supply thermocouple wire and complete thermocouples. Yet, a wide range of industries in various supply chains (automotive, chemical, consumer goods, HVAC, medical, metals, aerospace, petrochemical, plastics, and others) are major consumers of thermocouples. Such industries are very difficult to include in impact studies because the component supported by NIST infrastructure is a small portion of the downstream products that these industries produce and sell. Thermocouples are typically one of dozens, even hundreds, of components comprising products produced by the next level in the supply chain. These user industries are frequently unwilling to allocate time to

responding to a survey and, in fact, often cannot easily estimate the quantitative economic impact.

Support for thermocouple companies is an example of many NIST programs, which target single small producer goods industries. These industries serve many (often much larger) industries that comprise subsequent levels in supply chains. In such cases, pricing power is usually on the side of the larger purchasing industries. Thus, the majority of the economic impacts accrue to these larger users, which constitute the demand side of the markets addressed by the government infrastructure program. In contrast, the economic benefits to the supplier industry (thermocouples, in the example above), while significant, represent only a fraction of the total economic benefits delivered. These relatively small supplying industries are nevertheless critical to the entire supply chain because the key components provided affect the performance, quality, and reliability of downstream products and services or the efficiency of manufacturing processes. Unfortunately, measuring the impact on the single supplier industry is much easier than obtaining impact information from the several user industries.

Such situations are also frustrating from a public policy point of view. Each user industry may benefit only a modest amount from the government-supplied infrastructure, but collectively the economic benefits to these downstream industries are substantial, as the NIST impact studies indicate. This phenomenon (wide use in small amounts) is a characteristic of public goods, including technical infrastructure, and constitutes a primary justification for government involvement. Ironically, the more "public" (widely used) the infratechnology, the more difficult is the economic impact assessment. Thus, the program management lesson here is to exert maximum effort in the strategic planning stage to establish contacts with at least several of the user industries and to maintain these contacts throughout a project to both increase utilization of the project's results and to enhance economic impact estimation.

In contrast, in a study of NIST's Alternative Refrigerants Research Program (Shedlick et al [1998]), the results of NIST research were targeted at two supplier industries consisting of a relatively small number of large companies: manufacturers of refrigerants and manufacturers of refrigeration equipment. Here, identifying and surveying private sector users of a NIST database on refrigerant materials characteristics was relatively easy. The resulting economic impact estimates were therefore judged to be inclusive with respect to the impacted companies.

In many impact studies, the failure to measure economic benefits from more than one industry in the relevant supply chain may be the fault of the analysts conducting the study. However, low levels of industry interaction by the R&D agency not only reduce economic impact, as described in the previous section, but inhibit access by the analyst to that industry, which resulted in lower levels of *measured* economic impact. For example, in one of the three case studies conducted in the economic impact assessment of NIST's laser and fiberoptic calibration program, both NIST and the analyst likely contributed to an underestimate. NIST was not able to provide any leads or contacts in the lithography equipment industry. The analyst therefore constrained the data collection to the previous level in the supply chain (the source and detector manufacturers), claiming that the share of the economic benefits believed to have accrued to equipment manufacturers from

calibration infrastructure's major impact—prevention of out-of-spec equipment being produced and delivered to customers—could not be quantified.

In summary, microeconomic impact studies are limited to impact data obtained directly from industry by survey. Such an approach is usually the only method for acquiring quality impact estimates, as published data are almost always too aggregated to be of use. In general, the analyst is limited to collecting impact data from industries that are directly affected by the R&D agency; that is, industries that have direct contact with that agency and receive technical infrastructure support directly from it. Such industries tend to be both more willing to respond to surveys and more capable of providing usable quantitative data. Unfortunately, indirectly benefiting industries often realize the largest economic gains for a variety of reasons. Identifying such situations and establishing industry contacts would greatly increase measured economic benefits.

4.4.3. Managing Research Programs Over the Technology Life Cycle. Along with the supply chain, the technology life cycle is one of the two most important units of analysis for R&D policy (Tassey [forthcoming]). A number of the NIST economic impact studies strongly indicate the need to manage research programs and the accompanying set of projects in relation to the relevant industry's technology life cycle. Many government research programs attempt to supply multiple elements of an industrial technology, such as generic technology, a set of infratechnologies (measurement methods, implementation services such as SRMs, calibrations, and technical consulting), and standardization. The provision of each of these elements hopefully is optimally timed with the phases of the evolution of the technology as driven by private investment. The timing and longevity of each project within a research program are extremely important factors in determining the effectiveness of program design and management and therefore should be part of an economic impact study.

Allocation of resources among the major categories of projects (generic technologies, infratechnologies, and technology transfer activities) tends to progress sequentially over the technology life cycle. For example, for infratechnologies, methods are developed first followed by supporting services, although projects in each category often overlap and even recycle when necessary. Several of the NIST infratechnology research programs delivered large economic impacts with initial projects that produced new measurement methods, but the net benefits declined in later projects that provided supporting services. However, such a pattern does not mean that technology infrastructure services should be expected to under perform method development. In fact, they are usually essential for transferring the method to industry and then supporting its efficient use. That is, they enable the high benefits from the measurement method to be realized (hence, the conclusion that the impact of a program should be greater than the sum of its projects, if the latter were conducted in isolation). The apparent differential in net benefits arises when projects within a research program are analyzed in isolation, which usually ignores synergistic effects among projects.

NIST's antenna metrology program is an example of how supporting services are essential for transferring much of the potential economic benefits embodied in a method or interface specification. In fact, the technology transfer process was found to be surprisingly long and difficult in this case, because industry initially did not appreciate the economic value of the new near-field antenna measurement method. Eventually, the

NIST-developed method did become the industry standard.[58] The program management requirement is that calibrations and SRMs be tied to methods that are compatible with current industrial technology trajectories.

In well-integrated infratechnology programs, successive projects (calibrations, SRMs, and technical consulting services, along with general support for standardization) can have substantial economic impact for some time. However, as the technology trajectory being supported begins to decline, a new research program must be initiated in a time frame that provides the fundamental infratechnologies (such as measurement methods) needed by the new technology. Several impact studies indicate a tendency to extend the later phases of a program's activity (services and standardization, in particular) beyond what seems desirable relative to the evolution of the relevant industry's technology life cycle. Extending these services late into the technology life cycle not only results in a steady decline in the rate of net benefits, but seems to inhibit strategic planning for the next life cycle.[59]

That is, as the current technology life cycle matures, investment in research to develop infratechnologies (methods, data, etc.) for the next life cycle should be initiated. Otherwise, as happened in several programs studied, the next technology life cycle emerges and NIST is still overly focused on serving the previous one—primarily with services based on the first generation method. The strategic planning lesson is that an exit strategy coupled with a next generation investment strategy should be a prominent element of program planning and project management.

Another lesson from these impact studies relates to the utility to management of scenario planning. NIST's Alternative Refrigerants Research Program provides an example. In 1982, NIST initiated efforts to characterize the chemical properties of alternative refrigerants and determine how each of these new refrigerants performs when mixed with other refrigerants. This research was begun in anticipation of an international agreement to phase out the use of CFCs (chlorofluorohydrocarbons). Such an agreement (the Montreal Protocol) was reached in 1987. Delivery to industry by NIST of comprehensive and reliable data, along with analytical models for assessing the refrigerants' properties for equipment design, accelerated industrial R&D and reduced the cost of this R&D. By 1989, significant economic benefits from the NIST research were realized by U.S. refrigerant manufacturers from these R&D efficiency gains. Further benefits were realized by heating and cooling equipment manufacturers into the mid-1990s.

Finally, many believe that as technologies mature the opportunities for substantial economic impact from technical infrastructure support declines. Consequently, strategic

---

[58] An economic impact study of this program was attempted, but although most observers believed the economic impacts were substantial, the study was not completed due to data quality problems resulting from the extremely long time series of net benefits (almost 30 years) constructed by the analyst.

[59] Private companies have the same problem. Those firms that survive industry shakeouts in the early phases of a technology life cycle achieve substantial market shares. The resulting "cash cows" are hard to give up on and thus investment in the next generation of the technology is put off. As a result, new firms with new technologies eventually take market shares from those happily milking the defender technology.

planning in R&D agencies tends to emphasize high-growth industries as targets for funding. However, as the study of one NIST program (SRMs for sulfur in fossil fuels) showed, very large economic impacts can be obtained from investments in infratechnologies and associated standards for mature but large industries. After all, high-growth industries are important because of the implication that they will become large at some point in the future. Therefore, existing large industries should not be ignored when allocating resources among different economic sectors.

In summary, effective "technology life cycle management" should not result in a trade-off between resources to support high-growth industries and emerging technologies on the one hand and large but still important industries on the other. Presumably, mature industries are fairly well supported by existing government research budgets and reprogramming is undertaken as these industries actually decline in economic importance or their technology becomes static (which means they will decline). Studies of the need for infrastructure support of emerging technologies should result in new funds, rather than force unwarranted reprogramming from competitive but mature industries.

4.4.4. Net Economic Benefits from the Transition *between* Technology Life Cycles. In addition to managing the content of an infrastructure research program throughout a technology's life cycle, long-term delivery of economic benefits also requires identification of the transition between life cycles and the initiation of strategic planning to prepare for that transition. Adaptation of a government research program's content within a technology life cycle is demanding enough, but managing transitions between life cycles can be even more difficult (Tassey [forthcoming]).

An economic impact study of NIST's Cholesterol Standards Program (Leech [2000]) offers an example of how a transition to a new technology life cycle can shift both the industrial technology trajectories and the market structures that deliver or use the resulting technologies. The cholesterol impact study examined the last 14 years of a more than 30-year NIST program to support cholesterol measurement. Going back any further in time was not feasible due to data availability problems. This situation was especially unfortunate in the case of cholesterol measurement research, as qualitative analysis of the entire NIST Program's history indicated that its impact was greater in the first half of the time period.

Two reasons seem to explain the decline in impact: a major change in the generic measurement technology and, subsequently, a significant change in the industry's structure. The "wet chemistry" (strong acid) technology in the first part of this 30-year period required substantial and sophisticated analytical steps by clinical laboratories, which relied on NIST SRMs to achieve traceability. However, by the early 1990s, the development of enzymatic reagent methods ("dry chemistry") had created the potential to automate the analytical process and also allow the simultaneous analysis of multiple blood chemicals, including cholesterol. These technological trends coupled with cost reduction pressures created incentives to instrument manufacturers to provide automated analysis capability in the form of closed systems. In such systems, the instrument, reagents, calibrators, and controls are sold and assured only as a package or system.

This major shift in technology "de-skilled" the clinical laboratories—NIST's primary industry customers for its SRMs. With increasing automation, in which measurement

instruments and related diagnostic chemicals are sold as an integrated system, the wide use of NIST SRM's declined. Instrument and chemicals manufacturers (often the same company) took over responsibility for assuring accuracy from the clinical laboratories.

The lesson for strategic planning revealed by the economic impact study is that the government role of supplying a few primary standards at the apogee of a standards/calibrations pyramid is consistent with NIST's standards strategy, which emphasizes a progression toward increasing leverage through primary standards that drive a much larger secondary standards infrastructure. However, in this case, manufacturers have been certified against Center for Disease Control (CDC)'s reference method since 1982. They have therefore not relied directly on NIST's primary standards, instead depending on CDC-certified labs. The shift in the technical infrastructure supporting cholesterol measurement accuracy took place before the beginning of the 14-year study period selected by the analyst. This factor and the decline in economic impact often observed late in technology life cycles explain the muted economic impact relative to expectations.

None of this history suggests that NIST support was inappropriate in either amount or timing. In fact, the Institute's strategic planning recognized the decline in requirements for NIST support of this particular element ("analyte") of blood chemistry relative to other analytes and made programmatic adjustments. With current technology, instrument manufactures now provide clinical laboratories with the capability to simultaneously analyze 11 to 13 elements or compounds of interest to patients and their physicians. In recent years, NIST appears to have played a more critical role in assuring accuracy for a number of these other analytes. However, the study was started around the time this shift in programmatic direction was taking place, so the completed study appeared mismatched with the subsequent refocused program content. Today, an economic impact study of this program would be broader in scope to include at least several of these other analytes.

In summary, measurement methods, interface specifications, and other infratechnologies have a large impact early in the technology life cycle. Supporting infrastructure and services, such as SRMs and calibrations, increase this impact on a unit time basis and extend it over longer time periods. Equally important, supporting infrastructure is essential for transferring the infratechnology (method, interface protocol, etc.) to industry. Thus, the total economic impact can be greater than the sum of individual projects undertaken in isolation from each other. The timing of government infrastructure research over the technology life cycle and the transition between life cycles is therefore extremely important in determining the magnitude of economic benefits delivered.[60]

4.4.5. Competing Technologies and the Initiation of Economic Impact Assessments. One of the most difficult strategic planning problems for the manager of a government research program is to decide how to allocate scarce resources among competing versions

[60] The life cycle transition problem is even more acute for generic technology funding programs like NIST's ATP. Here, the focus is on the transition itself. ATP helps fund radically new generic technologies to which industry is failing to allocate sufficient resources due to a number of market failure mechanisms (Tassey [1997, forthcoming]). The major ones are spillovers, high time discounting, and mismatches with existing research capabilities and market foci.

of a technology that frequently appear in the early phases of the technology life cycle. These competing technologies can coexist for some time and thus spread available funds over a wide range of projects.

Alternatively, a “bet” can be placed on one of the competing technologies and resources concentrated. Managers of supporting infratechnology research programs can make similar bets. Such strategies increase the probability of attaining threshold levels of generic technologies and infratechnologies, thereby facilitating the timely development and implementation of market applications by industry. The risk is obviously that the bet will be placed on the wrong technology.

Such a risk was demonstrated by a NIST project established to support standards for a transport technology for high-speed broadband networks. The technology was asynchronous transfer mode (ATM), which helped create backbones for networks in the United States and elsewhere. Immediately upon initiation of the study, the NIST contractor observed that competing backbone technologies were gaining market penetration at the expense of ATM. Although the assessment was that ATM might eventually succeed, insignificant commercialization had occurred at the time of initiation of the economic impact study. The study therefore would have had to have been converted into a prospective analysis in order to continue.

In addition, preliminary discussions with industry indicated that the consortium (ATM Forum) attempting to set interoperability standards seemed to be going ahead with standards development with or without NIST. For example, industry did not seek NIST assistance for testing support. Rather, the ATM Forum formed a testing group and issued a call for participants to develop testing standards on its own.

In contrast, in the same general area of information technology, NIST’s work in developing and implementing the Data Encryption Standard (DES) for electronic funds transfer had substantial impact, in large part because a technology—electronic banking—was emerging and the industry needed a standard for transaction security (Leech and Chinworth [2001]). Without this standard, suppliers of hardware and software for encryption would not have entered the market, nor would many users have adopted electronic funds transfer technology. Here the existence of competing technologies and their associated standards was not an issue. Rather, a need was clearly determined for a specific standard to enable multiple market applications of an established generic technology and the industry needed the technical expertise and third-party impartiality of NIST.

4.5. Program Impact Assessment

Although they can be analyzed separately, most government R&D projects are part of broader research programs. Such programs seek to achieve major advances in a set of related technology infrastructure over time.

Obviously, policy makers need to be able to assess the overall performance of research programs, as well as efficiency of individual projects. Two methodological options are (1) devise and estimate performance metrics for the entire program, or (2) conduct a series of individual project impact assessments and use the collective results to infer overall program efficiency. The first option might seem preferable because only one

study is undertaken. However, frequently only general questions about *average* impact can be constructed and asked at the broader program level. That is, relating an entire government program consisting of multiple projects to the ultimate market penetration of the broad generic technology, which often is being pursued and/or absorbed by a large and heterogeneous set of firms, may only be accomplished at a general and largely qualitative level. Moreover, the multiple research projects aimed at each of several elements of the overall generic technology are likely to be somewhat out of phase with each other, which will compound measurement and attribution problems for the analyst.

Alternatively, conducting individual impact assessments of a subset of the program's projects provides more focused but more accurate economic impact data and also allow at least some degree of extrapolation to the program level. NIST's ATP adopted this approach to assess the economic impacts of its Component-Based Software Development (CBSD) projects (White and Gallaher [2002]). CBSD enables the re-use of software code and thereby increases the reliability of new software programs and facilitates the interoperability of individual applications and within enterprise-wide applications. However, this software development approach has suffered from lack of automated development tools for building and reusing components, as well as specification of component interface protocols.

ATP's CBSD Program funded 24 projects between 1994 and 1997 to fill these development tool and technical infrastructure gaps. As of 2002, 16 of those had been completed and 3 remained in progress; 5 failed for various reasons. ATP invested $41 million in the 19 projects and industry contributed $25 million. ATP surveys determined that as of 2002, industry had invested an additional $27 million in follow-on R&D. For the program impact assessment, the analysts selected 8 projects for study and collected cost and impact data for each project. The economic impact estimates for the combined projects are presented in Table 7.

Conducting such a study of multiple projects is demanding methodologically. One requirement for a "focus program" is to provide infrastructure technology elements to several levels in the relevant supply chain.[61] Thus, the analyst must identify the levels (industries) in the supply chain that are targets of the government research program and conduct separate surveys of each industry. The resulting data must then be aggregated into a single program-level impact estimate.

In this study, the analysts constructed 11 hypotheses about the nature of the government's role in the context of a conceptual framework of the software development process and underinvestment phenomena. Surprisingly, a consensus framework for software development does not exist, causing the analysts to synthesize one from the available literature. Eight projects were selected for quantitative performance. The quantitative analysis was coupled with qualitative analysis of the ATP role to produce the overall economic impact assessment. This report clearly demonstrates the added layer of complexity encountered in program-level analyses.

[61] In the mid-1990s, ATP initiated the "focus program" concept, modeled after DARPA's approach to funding emerging technologies. The central philosophy is that a broad (supply-chain wide) technology infrastructure (both generic technologies and infratechnologies) need to be funded to establish a threshold-level technology platform for subsequent private sector R&D investment decision making.

Table 7

Economic Impact Assessment of a Portfolio of Projects

ATP's Component-Based Software Development Program
(2000 dollars with projections for 2001 through 2004)

Metric	Estimate
Total Investment Costs	$119 million
Net Present Value	$840 million
Benefit-Cost Ratio	10.5
Social Rate of Return	80%
Total Producer Surplus	$538 million
Total Consumer Surplus	$1,129 million

Source: RTI International

4.6. Summary of Retrospective Impact Study Methodology. Once a screening process has identified a program or project for study, the major steps in designing, conducting, and interpreting the resulting analysis for stakeholders are

(1) Construct a taxonomy of the relevant technology, which disaggregates the technology into its major elements.

(2) Map this taxonomy onto the industry structure and competitive dynamics associated with development and delivery of the technology.

(3) Develop testable hypotheses that represent the relationships among technology, strategy, and economic trends.

(4) Utilize this framework to select a set of qualitative and quantitative output and outcome metrics.

(5) Select measures that summarize the metrics and are intelligible to stakeholders and other target audiences.

(6) Develop and implement a data collection plan, emphasizing primary data collection from impacted industries.

(7) In analyzing results, make careful determination of degree, if any, to which results can be extrapolated to other economic sectors.

(8) Write full report and then prepare several summaries of varying lengths for each target audience.

Successfully applying this generic methodology (summarized in Fig. 5) requires a multidisciplinary project analysis team and sufficient resources to ensure adequate primary data collection. Emphasis on the industries directly affected by the government

Figure 5

Steps in the Economic Assessment of Government R&D Projects

Technology and Industry Overview	Economic Analysis Framework	Data Collection	Economic Analysis	Results
▪ Technology trajectories ▪ Industry structure and competitive dynamics ▪ Technology Infrastructure (TI) trajectories ▪ TI market failures ▪ TI roles (project objectives or potential government roles)	▪ Identification of economic functions of TI ▪ Development of hypothesized economic impacts of TI or lack of TI ▪ Construction of economic metrics and selection of related measures ▪ Selection of study period	▪ Determination of industry populations ▪ Selection of survey methods ▪ Industry introductions ▪ Pretest survey ▪ Conduct industry surveys ▪ Collect cost data (retrospective studies)	▪ Quantitative analysis ▪ Qualitative analysis	▪ Draft report ▪ Final report ▪ Oral briefings ▪ Summaries for different audiences ▪ Publications

R&D being studied means that the estimated net benefits are derived directly from primary (industry) data sources and therefore have more credibility than indirectly estimated metrics. Focusing on corporate finance metrics not only is compatible with the project focus, but produces net benefit estimates that are readily understood by industry stakeholders.

5.0 Prospective Economic Impact Studies

5.1. Roles

For the most part, major and sustained allocations of R&D funds by R&D agencies have been made to specific technological areas without systematic quantitative analysis. Such analysis should include the relative costs of different investment barriers across these technologies and the relative costs within technologies of specific technical barriers that occur over the technology life cycle. Recognition of the weaknesses in the current resource allocation process has prompted demands for more systematic and analytical approaches to program development.[62]

Prospective or strategic planning information takes a variety of forms. These include results of workshops and conferences, customer needs surveys, and analyses based on

[62] For example, in 2002 OMB issued “investment criteria” for basic science and applied R&D research programs.

substantial data from primary sources. Such information is used either independently or integrated into a formal planning process. Information from ad hoc exercises such as workshops can identify technical barriers, including infrastructure needs and roughly rank them. Customer needs surveys can provide similar information with a more consistent structure and a somewhat greater degree of industry coverage, but with less detail. Both approaches, however, provide little or no quantitative information sufficient to rank barriers and allocate of resources among alternative technologies.

Effective strategic planning for government R&D programs requires appropriate combinations of qualitative and quantitative analysis, just as companies apply resource allocation tools based on a mix of the two. The mix varies depending on the stage of the R&D process being considered. For government research institutes and laboratories charged with providing technical infrastructure support to industry, an important first step is environmental scans across technologies to assess current and expected trajectories.[63] From such assessments, specific technologies can be ranked with respect to both economic importance and significant technology infrastructure problems. The second step is then to conduct in depth studies of the top-ranked candidates. The following example of software testing exemplifies the analytical methodology and level of effort required to estimate the aggregate and specific costs associated with perceived underinvestment in infrastructure for an important technology.

5.2. Software Testing Infrastructure

An excellent example of estimating the cost and hence potential economic benefits of technology infrastructure (in this case, infratechnology) is a study performed by RTI International [2002] for NIST of the costs to the economy of inadequate software testing. The working hypothesis was that underinvestment in software testing R&D was causing significant economic losses for the software supply chain and major using industries. The study is summarized below to illustrate the methodological steps required to not only achieve the desired economic analysis but to organize and present the results in way conducive to utilization by policy makers and industry stakeholders.

5.2.1. Software Testing as a Policy Problem. Few products of any type are shipped with such high levels of defects as is software. The media is full of reports of the catastrophic impacts of software failure, but these high-profile incidents are only the tip of a pervasive pattern of failure that software developers and users both agree is causing substantial economic loss.[64]

Anecdotal evidence points to testing inadequacies leading to embedded errors or "bugs" as the culprit. In fact, the process of identifying and correcting defects during the

---

[63] Large technology-based companies conduct scans. However, government R&D agencies must emphasize this step to an even greater degree because their potential research portfolios cover large economic sectors. In fact, in some cases, research portfolios span the entire economy (e.g., NIST in the United States and the Framework Program in Europe).

[64] The final consumer obviously incurs huge costs due to faulty software. Considerable lost productivity and income result from reduced efficiency or actual suspension of economic activity. However, these ultimate costs (for example, shutting down a manufacturing plant or a stock exchange) were not included in the study due to the difficulty in defining and estimating them.

software development process was estimated in the NIST study to represent approximately 80 percent of development costs. Moreover, complexity is increasing, which not only leads to more errors but makes error detection and removal more difficult. The size of software products is no longer measured in thousands of lines of code, but in millions. This greater complexity along with a decreasing average product life expectancy has increased the economic costs of errors. Over half of software bugs are currently not found until “downstream” in the development process, leading to significantly greater economic costs.[65]

Because testing methods and associated standards have an infrastructure character, the issue was deemed important for R&D policy analysis. In fact, the general problem was asserted by industry observers to be a lack of testing infrastructure that individual companies (developers and users alike) and even groups of companies are having difficulty providing. Thus, a study to characterize software errors and their nature and then to estimate their cost was in order.

In studies aimed at informing policy makers, the analysis must accomplish three steps: (1) carefully describe and characterize the technical nature of the barriers causing economic loss, so that an effective R&D program can eventually be designed (if warranted by estimates of the economic losses), (2) estimate the economic losses being incurred by industry, preferably by major subcategory of the infrastructure under study, and (3) demonstrate why the private sector cannot efficiently remove the technical barriers causing the loss.

Because an economic impact study of a specific technology is not feasible without extensive use of primary data sources and because such data collection is relatively expensive, critical decisions have to be made with respect to the scope and depth of the economic analysis. How this issue is resolved will determine levels of interest by stakeholders in the study results and the credibility of the results with respect to influencing policy responses. Selection of specific industries within the supply chains relevant to the technology under study will affect efficiency of data collection, as industries respond differently to requests for study participation. Moreover, the ability to eventually extrapolate case study results to the national economy level is strongly correlated with the scope and representative nature of industries studied and the degree of uniformity of software testing across areas of application.

In the NIST study, inadequate testing was defined as failure to identify and remove software errors in real time.[66] Two broad case studies were chosen for analysis. Transportation equipment (automotive and aerospace) was selected in the manufacturing sector because of its importance to the economy and because the relevant supply chains are typical of highly automated (computerized) manufacturing. Both the individual manufacturing processes and the increased vertical integration among the industries

[65] Industry now spends more than $1 billion dollars per year on software testing tools and this expenditure is projected to reach $2.6 billion by 2004. Approximately 302,000 workers are engaged in testing and debugging activities, which represents approximately one-fourth of all computer programmers and software engineers.

[66] Note that this focus on *detection and efficient removal* is distinctly different from the alternative topic of reducing the *number* of software errors.

making up these supply chains are highly software dependent. Financial services was chosen as a case study in the service sector, although this study also included the router and switch manufacturers whose highly computerized equipment and associated sophisticated software provide the essential backbone for financial services.

5.2.2. Selection and Construction of Metrics. Cost differentials constitute the primary impact variable in infratechnology assessments, but such differentials can be structured in several different ways. The process is to identify the product/service attributes of interest and then to construct metrics, which facilitate primary data collection. Metric development in the software testing study began with the International Standards Organization (ISO) list of six characteristics of software quality (obviously affected by errors) and an IEEE list of potential technical metrics. Selection from generic lists of metrics is based on the degree to which particular quality attributes are present and an understanding of what constitutes a good metric. For software testing, IEEE provides five "validity measures" to determine which metrics are most effective for the particular area of development or analysis. These technical metrics drive cost estimation.

Whatever economic metrics are eventually selected, they must match as closely as possible the assessed ability of the industries in the case studies to respond to data collection efforts. In the case of software testing, the lack of quality metrics (or, at least the lack of an industry consensus on what those quality metrics should be) leads most companies to simply count the number of defects that emerge when testing occurs. This situation significantly inhibits cost impact estimation.[67]

This situation led to a decision to define bottom-line cost metrics that represented aggregates of costs associated with the general structure of testing, rather than define costs associated with specific testing methods or activities. For example, costs were segregated by "pre-release labor costs", "hardware costs", "software costs", "external testing costs", and "after sales service costs". The selected approach was more compatible with how company cost data might be kept or compiled.

Industry respondents first were asked to provide data for the designated cost categories under current practice (baseline scenario) in which delays occur between the introduction of errors and the identification/removal of those errors. To achieve the desired estimates of cost differentials associated with inadequate testing, a counterfactual was constructed in which the respondents were asked to estimate cost reductions for hypothetical situations in which errors were detected and removed in the same phase of software development. This last step is important because it potentially allows the estimation of the three financial ratios (impact measures) described in Section 2.

5.2.3. Data Collection and Analysis. For the transportation equipment case study, data were collected from software developers (CAD/CAM/CAE and product data management vendors) and from users (primarily automotive and aerospace companies). For the service sector case study, financial services were analyzed with data collected

[67] Research Triangle Institute [2002, p. 1–7]. Few organizations seem to use advanced testing techniques, such as forecasting field reliability based on test data and calculating defect density to benchmark the quality of their product against others.

again from software developers (routers and switches, financial electronic data interchange, and clearinghouse) and from users (banks and credit unions).[68]

The study concluded that the path to higher software quality is significantly improved software testing, as opposed to more testing.[69] However, testing methods have a strong infrastructure character because generally accepted (standardized) approaches must be used to assure buyers that higher quality levels have in fact been achieved. This requirement for common use (standards) coupled with shortening technology life cycles and subsequent pressures to get new generations of software into the market place ahead of competition lead to substantial underinvestment in the infratechnologies underlying software testing.

As indicated in Table 8, without standardized and comprehensive advanced testing infrastructure, the estimated annual cost to these two major industry groups from inadequate software infrastructure is \$5.85 billion. Similarities across industries with respect to software development and use and, in particular, software testing labor costs allowed a projection of the cost to the entire U.S. economy. Using the per-employee impacts for the two case studies, an extrapolation to other manufacturing and service industries yields an estimate of \$59.5 billion as the annual cost of inadequate software testing infrastructure.

The results of this study would seem to easily justify undertaking it. However, such *a priori* motivations typically are based on anecdotal evidence, which is often far from accurate. In this case, considerable information was available to support the proposition that significant economic costs were being incurred. Additional anecdotal evidence indicated that advanced testing methods are not readily available, fully proven, accompanied by test suites, and then accepted as industry standards. Standardized testing tools, suites, scripts, reference data, reference implementations, and metrics that have undergone a rigorous certification process were thought to offer the potential for a large impact on the testing inadequacies currently plaguing software markets.

However, successfully demonstrating that significant economic costs are being incurred creates the problem of determining the proper government/industry response. Ideally, if all software bugs could be identified and removed in the phase of development in which they occur (i.e., in real time), the combined economic benefits to the two

[68] The basic forms of data collection instruments and related techniques have already been discussed in Section 3.3.

[69] In all, the study estimated that 80 percent of software development costs are already allocated to testing activities, so simply doing more testing does not seem to be a high-payoff strategy. Rather, improving the *efficiency* of the testing infrastructure by developing *better* test methods, which industry can adopt as standards, appears to be the logical direction of response.

Table 8

Economic Analysis of Software Testing Infrastructure ($billions)

	Cost of Inadequate Software Testing Infrastructure			**Potential Cost Reductions from Infrastructure Improvements**		
Supply Chain Segment	**U.S. Economy**	**Transportation Equipment**	**Financial Services**	**U.S. Economy**	**Transportation Equipment**	**Financial Services**
Software Developers	21.2 (36%)	0.37	2.34	10.6 (47%)	0.16	1.20
Software Users	38.3 (64%)	1.47	1.01	11.7 (53%)	0.43	0.31
Total	59.5 (100%)	1.84	3.34	22.2 (100%)	0.59	1.51

Notes: (1) Transportation Equipment primarily consists of automotive and aerospace industries; (2) Software developers include CAD/CAM/CAE and PDM in transportation equipment and FEDI (Financial Electronic Data Interchange), clearinghouse software, and software for routers and switches in financial services.

industry groups and to the economy would be $5.85 billion and $59.5 billion, respectively. Completely removing complex technical infrastructure barriers is not possible. Realizing that such a “perfect infrastructure” is not attainable, the analysts asked industry experts during the data collection phase for estimates of a plausible reduction in delayed identification and removal of software errors. From this information, a “feasible improved infrastructure” scenario was constructed. This improved infrastructure scenario was estimated to result in a combined annual benefit of $2.10 billion to the two industry groups studied and $22.2 billion to the entire U.S. economy.

The study’s estimate of $22 billion in potential economic cost reduction from a program of reasonable technical goals and moderate financial cost indicates the potential gross benefits of a government policy response. A detailed analysis of the technical and economic barriers preventing provision of adequate software infrastructure provides the basis for an estimate of the R&D program required to remove this portion of the estimated total costs of the and hence provides an estimate of the underinvestment in R&D.

The requirement for a government role derives from the public good nature of the needed testing infrastructure. Such infrastructure, embedded in standards, has significant value only if commonly and uniformly used. Common use connotes common ownership among competing suppliers and their customers and hence provides a disincentive for individual companies to make the needed investments. Consortia of firms often form to promulgate standards, but such collaborations often lack the technical expertise and the resources to provide the infratechnology basis for those standards.

5.2.4. Methodological Implications. This case study exemplifies the multidisciplinary approach required to design and conduct a prospective analysis of a particular technology infrastructure, and then to interpret the results in a way that is digestible by policy makers and other stakeholders. Rigorous design at the microeconomic level of technical metrics and appropriate measures of impact coupled with careful, systematic primary data collection is the only way to provide credible, decision-relevant results. Metrics must be seen as correctly capturing the complexity of the overall technology and, in particular, its infrastructural elements. The selected impact measures must be widely accepted and deemed appropriate for the level of aggregation of the economic activity being analyzed.

As more studies of the above type are conducted, methodologies will be refined and the results will be more accepted by decision makers. Specifically, because of their complexity, such studies must be conducted with sufficient frequency to not only gain acceptance for the general methodology but to achieve a high level of acceptance by decision makers with respect to understanding and use of such studies’ results.

A sample of recent prospective economic studies conducted by NIST is shown in Table 9. The scope of these studies (industry coverage) varies, as does the type of metrics estimated. The first study in the table, an analysis of the costs of inadequate interoperability in the automotive supply chain, estimated an annual cost to that supply chain of $1 billion. Because this study collected data from one supply chain (actually a portion of that supply chain), the cost estimates were not extrapolated to the national level, as was done in the software testing study. Nevertheless, the quality of the analysis and the recognition by stakeholders that other manufacturing supply chains (aerospace,

for example) had similarly costly interoperability problems led Congress to pass the Enterprise Integration Act in 2002. This legislation authorized NIST to help industry develop infratechnologies and associated standards to reduce the costs of incompatible data file formats and thereby increase the efficiency of electronic data exchange among industries in a supply chain.

Table 9

Studies of Costs Due to Inadequate Infrastructure Investment

Focus of Study	Industries Covered	Infratechnologies Studied	Estimated Annual Costs
Interoperability costs	• Automotive supply chain	• Electronic product design data format standards	$1 billion
Deregulation	• Electric utilities	• Measurement methods • Systems monitoring & control techniques	$3.1–$6.5 billion
Software testing	• Transportation equipment • Financial services	• Testing tools and techniques	$60 billion
Medical testing	• Medical testing labs (one of 3 case studies) • Hospitals	• Measurement and quality assurance techniques	$0.4–$1.3 billion

The cost estimates of other studies in Table 9 also vary in scope and hence estimated costs. The estimated cost of inadequate medical testing is from the first of three planned case studies (blood tests for calcium). Obviously, when all three case studies are completed, the estimated cost will be considerably greater and likely draw more attention from stakeholders.

The study of the implications of deregulation of electricity generation was primarily an assessment of the technical issues arising from the effective creation of new markets (vertically integrated utilities now may only engage in one or two of the three basic stages of electricity services: generation, transmission, and distribution). These market interfaces require new standards to address transaction costs that are always incurred in market exchanges between industries and vertical supply chains. The study, however, did not emphasize quantitative analysis, so the estimated costs were appropriately characterized in the study as approximate.

6.0 Use of Economic Analysis by Policymakers

Economic analysis by decision makers in the science and technology policy arena is only beginning to be effectively utilized. Retrospective studies, if done in sufficient quantity to be deemed “representative” of an R&D agencies programs, can “validate the

business model" of that agency and even justify funding new programs. Such studies also give credibility to the predictions of future economic impacts by prospective studies. However, prospective studies are essential to definitively characterize and quantify existing technical barriers.

Economic studies of both types illuminate the factors affecting the nature and magnitudes of outcomes and thereby also provide a qualitative context for helping to understand the technological and market forces contributing to the estimated economic impacts. Thus, economic impact assessment should be an ongoing function in R&D agencies to generate a database sufficient to provide retrospective and prospective impact indicators for the technologies and industries targeted by an agency's mission and to help manage mission modification over time.

However, in depth microeconomic impact studies are fairly expensive to conduct, so only a limited number of R&D projects can be assessed in a given time period. As a result, these studies should be viewed as inputs to a broader performance measurement system. Such a system includes a set of generic performance metrics and measures common to projects within major programmatic areas. The general metrics and measures can be tracked on a regular and relatively inexpensive basis and thus are appropriate for meeting annual reporting requirements such as those under GPRA. Although not as project specific or empirically based as the metrics and measures produced by microeconomic studies, the microeconomic studies can leverage the more general metrics through extrapolation and validity testing effects.

More generally, performance metrics should be tailored to the type of technology infrastructure studied and the nature of the target audiences. Audiences include the S&T and economic policy arenas, the GPRA and executive branch efficiency processes, and program management and strategic planning staff within government R&D agencies.

7.0 Appendix

Microeconomic Analyses of NIST Infratechnology (Laboratory) Research

Industry: Project	NIST OU/Year	Output	Outcomes	Measures
Semiconductors: resistivity	EEEL/1981	test methods	increase productivity	SRR: 181% BCR: 37
Semiconductors: thermal conductivity	EEEL/1981	materials properties test methods	increase R&D efficiency lower transaction costs	SRR: 63% BCR: 5
Semiconductors: wire bonding	EEEL/1981	test methods	increase productivity increase R&D efficiency	SRR: 140% BCR: 12
Communications: electromagnetic interference	EEEL/1991	test methods	lower transaction costs	SRR: 266% BCR: 12.6
Semiconductors: electromigration	EEEL/1992	test methods	increase R&D efficiency transaction costs	SRR: 117%
Photonics: optical fiber	EEEL/1992	test methods (acceptance)	lower transaction costs	SRR: 423% BCR: 17.2
Automation: real-time control systems	MEL/1995	generic architecture	increase R&D efficiency	SRR: 149% BCR:
Energy: electric meter calibration	EEEL/1995	test methods (calibration)	lower transaction costs	SRR: 117% BCR: 12
Communications: ISDN	ITL/1995	interoperability standards	lower transaction costs	SRR: 156% BCR:
Computers: software conformance	ITL/1995	test methods (acceptance)	lower transaction costs	SRR: 41% BCR:
Photonics: spectral irradiance	Physics/1995	test method (calibration)	increase productivity lower transaction costs	SRR: 145% BCR: 13
Construction: building codes	BFRL/1996	technical basis for standards	energy conservation energy cost savings	SRR*: 57% BCR:
Construction: roofing shingles	BFRL/1996	materials properties	increased durability	SRR*: 90% BCR:

Industry: Project	NIST OU/Year	Output	Outcomes	Measures
Construction: fire safety evaluation system	BFRL/1996	technical basis for standards	lower compliance costs	SRR*: 35% BCR:
Automation: machine tool software error compensation	MEL/1996	quality control algorithm	increase R&D efficiency increase productivity	SRR: 99% BCR: 85
Materials: thermocouples	CSTL/1997	standard reference data (calibration)	lower transaction costs increase product quality	SRR: 32% BCR: 3
Pharmaceuticals: radiopharmaceuticals	Physics/1997	standard reference materials	increase product quality	SRR: 138% BCR: 97
Photonics*:* optical detector calibration	Physics/1997	standards and calibration services	increase productivity	SRR: 72% BCR: 3
Chemicals: alternative refrigerants	CSTL/1998	standard reference data	increase R&D efficiency Increase productivity	SRR: 433% BCR: 4 NPV: $5.6M
Materials: phase equilibria for advanced ceramics	MSEL/1998	standard reference data	increase R&D efficiency increase productivity	SRR: 33% BCR: 10
Semiconductors: software for design automation (IGBT semiconductors)	EEEL/1999	software model	increase R&D efficiency increase productivity	SRR: 76% BCR: 23 NPV: $10M
Pharmaceuticals: cholesterol measurement	CSTL/2000	measurement method standard reference materials	increase productivity reduce transaction costs	SRR: 154% BCR: 4.5 NPV: $3.5M
Photonics: laser and fiberoptic power and energy calibration	EEEL/2000	calibrations	increase productivity reduce transaction costs	SRR: 43%–136% BCR: 3–11 NPV: $48M
Chemicals: SRMs for sulfur in fossil fuels	CSTL/2000	standard reference materials	Increase productivity reduce transaction costs	SRR: 1,056% BCR: 113 NPV: $409M

Industry: Project	NIST OU/Year	Output	Outcomes	Measures
Electronics: Josephson voltage standard	EEEL/2001	standard reference materials	increase R&D efficiency increase productivity enable new markets	SRR: 877 BCR: 5 NPV: $18M
Quality: BNQA	MBQA	technical information	increase productivity	BCR: 207 NPV: $25B[70]
Communications: security (data encryption standard)	ITL/2001	standards; conformance test methods/services	Increase R&D efficiency enable new markets	SRR: 270 BCR: 58-145 NPV: $345M–$1.2B
Communications: security (role-based access control)	ITL/2001	generic technology reference models	enable new markets increase R&D efficiency	SRR: 62% BCR: 109 NPV: $292M
Chemicals: National Traceable Reference Materials Program (NTRM)	CSTL/2002	reference data; calibration services	increase efficiency of regulatory compliance (content & production efficiencies for standards)	SRR: 221% BCR: 21 NPV: $49M
Manufacturing: standards for product data exchange (STEP)	MEL/2002	standards development; conformance test methods/services	increase quality and assimilation of standards; accelerate standards development	SRR: 32% BCR: 8 NPV: $180M
SRR=social (internal) rate of return SRR*=social (implied) rate of return BCR=benefit-cost ratio NPV=net present value				

[70] Extrapolated to the national (economy-wide) level. This extrapolation distinguishes this study from the others, which only estimated benefits for the initial market or markets directly affected by the NIST program (i.e., only markets for which primary data sources could be accessed).

8.0 References

Au, T. and T.P. Au [1992], *Engineering Economics for Capital Investment Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Chase Econometrics, Inc. [1975], *The Economic Impacts of NASA R&D Spending*. Bala Cynwyd, PA..

CONSAD [1997], Advanced Technology Program: *The Development of Advanced Technologies and Systems for Controlling Dimensional Variation in Automobile Body Manufacturing*. Gaithersburg, MD: National Institute of Standards and Technology.

Fogel, Robert William [1962], "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings", *Journal of Economic History* 22:2 (June): 163–97.

Frederick, S., G. Loewenstein and T. O'Donoghue [2002], "Time Discounting and Time Preference: A Critical Review", Journal of Economic Literature 40 (June): 351–401.

Gallaher, M. and S. Martin [1999], *Benefit Analysis of IGBT Power Device Simulation Modeling* (Planning Report 99–3). Gaithersburg, MD: National Institute of Standards and Technology.

Gallaher, M., A. O'Conner, and B. Kropp [2002], *The Economic Impact of Role-Based Access Control* (NIST Planning Report 02–1). Gaithersburg, MD: National Institute of Standards and Technology.

Graham, J. and C. Harvey [2001], The Theory and Practice of Corporate Finance: Evidence from the Field", *Journal of Financial Economics* 60 (May/June): 187–243.

Griliches, Zvi [1975], "Comments on CEA: The Economic Impact of NASA R&D Spending". Harvard University, unpublished paper.

Griliches, Zvi [1977], "Economic Problems of Measuring Returns on Research" in *Towards A Metric of Science: The Advent of Science Indicators,* Vehuda Elkana, ed. New York: John Wiley & Sons.

Griliches, Zvi [1979], "Issues in Assessing the Contribution of Research and Development to Productivity Growth", *Bell Journal of Economics* 10 (Spring).

Griliches, Zvi [1988], "Productivity Puzzles and R&D: Another Non-explanation," *Journal of Economic Perspectives* 2: 9–21.

Griliches, Zvi [1995],

Kostoff, Ronald [2001], "Physics Research Impact Assessment", *Physics World*.

Link, A. and J. Scott [1998], *Public Accountability: Evaluating Technology-Based Institutions*. Norwell: Kluwer.

Leech, David [2000], *The Economic Impacts of NIST Cholesterol Standards Program* (NIST Planning Report 00–4). Gaithersburg, MD: National Institute of Standards and Technology.

Leech, D. and M. Chinworth [2001], *The Economic Impacts of NIST's Data Encryption Standard (DES) Program* (NIST Planning Report 01–2). Gaithersburg, MD: National Institute of Standards and Technology.

Link, Albert [1992], *Economic Impacts of NIST-Supported Standards for the U.S. Optical Fiber Industry*. Gaithersburg, MD: National Institute of Standards and Technology, Electronics and Electrical Engineering Laboratory.

Link, Albert [1995], *An Evaluation of the Economic Impacts Associated with the NIST Power and Energy Calibration Services* (NISTIR 5565). Gaithersburg, MD: National Institute of Standards and Technology.

Mansfield, E., J. Rapport, A Romeo, S Wagner, and G. Beardsley [1977], "Social and Private Rates of Return from Industrial Innovations", *Quarterly Journal of Economics* 91(May): 221–240.

Martin, S., M. Gallaher, and A. O'Conner [2000], *Economic Impact of Standard Reference Materials for Sulfur in Fossil Fuels* (NIST Planning Report 00-1). Gaithersburg, MD: National Institute of Standards and Technology.

Marx, M., J. Scott, and S. Fry [2000], *Economic Impact Assessment: NIST-EEEL Laser and Fiberoptic Power and Energy Calibration Services* (Planning Report 00-3). Gaithersburg, MD: National Institute of Standards and Technology.

Nadiri, Ishaq [1993], "Innovations and Technological Spillovers" (NBER Working Paper No. 4423). New York: National Bureau of Economic Research (August).

OECD [2002], *OECD Information Technology Outlook: ICTs and the Information Economy 2002*. Paris: OECD.

Pressman, R. S. [1992], *Software Engineering: A Practitioner's Approach, Third Edition*. New York: McGraw-Hill.

Research Triangle Institute [2002], *The Economic Impacts of Inadequate Infrastructure for Software Testing* (NIST Planning Report 02-3). Gaithersburg, MD: National Institute of Standards and Technology (May).

Ross, S., R. Westerfield, and J. Jaffe [2002], *Corporate Finance*. New York: McGraw-Hill Irwin.

Ruegg, Rosalie [1999], Advanced Technology Program's Approach to Technology Diffusion (NISTR 6385). Gaithersburg, MD: National Institute of Standards and Technology (September).

Shedlick, M., A. Link, and J. Scott [1998], *Economic Assessment of the NIST Alternative Refrigerants Research Program*. (NIST Planning Report 98-1). Gaithersburg, MD: National Institute of Standards and Technology.

Tassey, Gregory [1996], *Rates of Return from Investments in Technology Infrastructure* (NIST Planning Report 96-3). Gaithersburg, MD: National Institute of Standards and Technology.

Tassey, Gregory [1997], *The Economics of R&D Policy*. Westport, CT: Quorum Books.

Tassey, Gregory [1999], "Lessons Learned about the Methodology of Economic Impact Studies: The NIST Experience", *Evaluation and Program Planning* 22: 113–119.

Tassey, Gregory [forthcoming], "Underinvestment in R&D-Intensive Industries: Composition Market Failure".

Tewksbury, J., M. Crandall, W. Crane [1980], "Measuring the Societal Benefits of Innovation", *Science* 209 (August): 658–662.

White, W. and M. Gallaher [2002], *Benefits and Costs of ATP Investments in Component-Based Software* (NIST GCR 02-834). Gaithersburg, MD: National Institute of Standards and Technology (November).